# PHYSICAL SCIENCES

## THE ROYAL INSTITUTION LIBRARY OF SCIENCE

## VOLUME 10

# THE ROYAL INSTITUTION
## LIBRARY OF SCIENCE

# PHYSICAL SCIENCES

## Volume 10

# PHYSICAL

# SCIENCES

## Volume 10

EDITED BY

SIR WILLIAM LAWRENCE BRAGG,
CH OBE MC FRS MA Cantab
Professor Emeritus and formerly Director of the
Royal Institution of Great Britain and Fullerian Professor of Chemistry

AND

PROFESSOR GEORGE PORTER,
FRS MA ScD Cantab
Director of the Royal Institution of Great Britain and
Fullerian Professor of Chemistry

ELSEVIER PUBLISHING COMPANY LTD

# CONTENTS

## 1934

## 1935

## 1936

## 1937

## 1938

## 1939

MAJOR CHARLES E. S. PHILLIPS, O.B.E., F.R.S.E.
*Secretary and Vice-President, in the Chair*

ALLAN FERGUSON. M.A., D.Sc.
*Assistant-Professor of Physics, East London College*

## SURFACE TENSION

I cannot more appropriately introduce the subject of to-night's discourse than by repeating a simple experiment made at this table by Lord Kelvin almost exactly forty-eight years ago, when, at the beginning of a discourse on capillarity,* he dipped his finger into a basin of water and drew the attention of the audience to the " force of attraction between the finger and the drop hanging from it, and between the matter on the two sides of any horizontal plane you like to imagine through the hanging water."

Indeed, our subject is very closely associated with the comprehensive topic of cohesion, a topic which attempts an answer—not to the problem, *why* we are here to-night, which is a matter for the theologians to discuss—but to the problem, *how* we are here, in our present habits, and not as a chance medley of unattracting atoms.  To attempt to answer such a query takes us very far towards the fundamentals of atomic structure and behaviour, but the problem, as we envisage it in the light of to-day's theories, still bears a strong resemblance to the problem as it was posed, and answered, by van der Waals.  How comes it that,

* Proc. Roy. Inst.   Vol. XI.   p. 483.   1886

if material particles attract each other, the whole structure of the universe does not collapse under these attractions ? We can formulate an answer to the question if we take into account the thermal motions of the particles. " In nature it is cohesion between atoms which tends to produce condensation and solidification, and temperature which tends to produce dissociation . . . Temperature is a manifestation of kinetic energy and cohesion of potential energy and the interplay of these two forms of energy is responsible for many of the observed physical properties of matter."[*] Cohesion and temperature—these, then, are the protagonists who play out the drama. The main change of view during the last generation depends on the change which has taken place in our concept of the ultimate material particle and the structure of the atom. The billiard ball atom of the nineteenth century physicist served its purpose well, and subserves a useful function to-day. It is not surprising that so naive an extrapolation of our large scale processes should ultimately break down—the surprising matter is that the extrapolation should have proved so brilliantly successful as it has done. Such a type of atom was assumed to have a definite size, and this notion of clear cut dimensions was not lost when it was found that the concept was not fine-grained enough to interpret successfully radio-active phenomena, and the nuclear atom displaced it. Here the concept of size was associated with the dimensions of the electronic orbits and, though the atom became a far more complex system, there was no haziness about the notion, apart from the difficulty of an exact determination of its dimensions.

[*] Lennard-Jones, Proc. Phys. Soc. Vol. XLIII. p. 461. 1931.

We have changed all that to-day. We cannot hope to locate our electrons precisely, and the definite orbit of the electron of the nuclear atom is now replaced by a *probability pattern*, the density of the pattern at any point measuring the probability of finding an electron there. Fortunately the uncertainty which we have introduced into our concepts has not seriously altered our views concerning the order of atomic magnitudes, inasmuch as the probability of finding an electron in any particular pattern becomes infinitesimal outside regions of the magnitude of about one hundred millionth of a centimetre. We may still, therefore, keep to an interpretation of the size of an atom which, despite a little cloudiness at the edges, does not differ materially from the interpretation based on the older concepts.

What are, then, the dimensions of the atoms and molecules with which we have to deal ? Any example that we may give to illustrate the meaning of such atomic magnitudes merely transforms an inconceivably small number into an inconceivably large one ; let it suffice to say that if we could curb the liveliness of the hydrogen atoms and lay them in order on a sixpence, it would take some eight hundred million years of unceasing work to cover the coin, if we laid our atomic bricks at the rate of one a second.

We shall, therefore, picture a liquid as a congeries of such particles in lively thermal motion, attracting and being attracted according to a law which we need not attempt to specify more closely than by saying that the attraction of any one molecule on its neighbours falls off very rapidly as the distance increases. If then we draw round

any one molecule, well in the interior of the liquid, a small sphere whose radius we may term the range of molecular attraction, the central molecule will experience no resultant force due to the attractions upon it of its neighbours within this sphere. It is otherwise if the central molecule is at a distance from the surface of the liquid which is less than the radius of this sphere. Part of the sphere is now outside the liquid, the molecules therein are missing, and do not contribute their share to the force on the molecule at the centre of the sphere, which molecule therefore experiences a force urging it into the liquid. To transport a molecule from the interior of the liquid into the liquid surface against such a force requires, therefore, the expenditure of *work ;* and inasmuch as the conveyance of molecules into the surface means an extension of the surface, to extend a liquid surface necessitates the expenditure of work. Stretch a sheet of rubber, and notice that we have to do work to extend the rubber surface. We must not press simple analogies too far, but it is legitimate to assume that the surface of a liquid behaves as if it were in a state of tension, and the tension in the surface across a line of unit length drawn in the surface is called *the surface tension* of the liquid. We may demonstrate its existence by forming a soap film on a circle of copper across which a loose thread of cotton has been tied. Destroy the film on one side of the thread, and the thread is pulled by the tension of the film on the other side into a very perfect arc of a circle. This tension differs for different substances, as we may shew by pouring a little alcohol on to a thin film of coloured water at the bottom of a glass dish, and noting how a clear space is formed in the middle of the dish

where the alcohol—the liquid of weaker tension—was poured in.

For any one liquid, the tension decreases with increasing temperature. We may shew this by scattering lycopodium on the surface of water in a dish, and bringing a heated bit over the centre of the dish. The greater tension of the cold water is clearly shewn by the way in which the central space is swept clear of the powder.* A film of oil covering a thin rod breaks up into a series of regularly spaced drops, illustrating the instability of a liquid cylinder whose length exceeds its circumference. A similar result may be illustrated for a liquid annulus by cutting a narrow circular groove on the lower surface of an iron disk, placing the disk on a horizontal glass plate in the projector and filling the groove with mercury. On lifting the disk, the ring of mercury breaks up at once, as you see, into a series of regularly spaced drops.

These drops, we note, take on an approximately spherical form, and the tendency to sphericity becomes more pronounced as the drops become smaller, and those forces, such as the weight of the drop, become increasingly unimportant in comparison with the surface forces. A spherical surface is that surface which, for a given volume, has a minimum area, and this tendency to the exposure of the smallest possible extent of surface illustrates the principle that a dynamical system

---

* Following Lord Kelvin's dictum that we know something about a quantity when we can make measurements thereon, it may not be amiss to state that the law connecting surface tension ($\gamma$) with temperature ($\theta$) is $\gamma = \gamma_0(1 - b\theta)^n$, where $b$ is very accurately the reciprocal of the critical temperature, and $n$, which varies only slightly from liquid to liquid, may be assumed to have a mean value of 1·2. This relation holds good for *unassociated* liquids.

tends to take up a position in which its potential energy is a minimum.

These few fundamental principles, consistently applied, will serve to elucidate a great many problems in which surface energy plays a dominant part, and I had originally intended to illustrate these principles by a number of experiments and measurements of surface tension. Certain kinematographic experiments have, however, given results of such interest that I propose to consider them at some length, and have relegated illustrations of methods of measurement to a display in the library. One interesting experiment may, perhaps, be permitted, which we owe to the ingenuity of Major Phillips. Two light vertical rods serve as supports. To the top of one is fixed a *circular* microscope cover slip with its plane horizontal ; the other support carries a horizontal *square* cover slip. Two other slips, one square, one circular, carry fastened to their upper surfaces long and light straw pointers. Lay these on the top of the fixed slips, and, clearly, you can spin them round as you will. But now place a drop of water between each of the pairs of slips, and endeavour to revolve the upper movable slips. The circular one moves freely over its circular fellow and remains with its pointer pointing in any direction which you choose to give to it. Displace the square one, and it snaps back into a perfectly definite position of equilibrium in which the upper slip is exactly congruent with the lower. It is a delightful experiment, and one is tempted to say of it, *mutatis mutandis*, what Praed said of the Vicar's sermons.

But we must hasten to the second part of the

discourse, which is concerned with a very commonplace phenomenon—that of the detachment of a drop of liquid from a vertical tube. Study the process at any slowly dripping tap, and you will see that, although the initial stages may very easily be examined, the final stages of detachment occur with a rapidity which makes it impossible for the eye to follow them. How may we slow the process ? Aniline and water are almost immiscible, and the density of aniline is but slightly greater than that of water. If, therefore, we form an aniline drop at the end of a vertical tube dipping into water, we may project the image of the drop on to a screen, and study the circumstances of its detachment with much greater ease.

Pitch is a queer substance ; to forces of short duration it behaves as a solid, leave it to itself and it will flow like a liquid. Here we have some pitch in a funnel, and you observe a very perfect drop of pitch pendent from the stem of the funnel. It seems static enough ; but although it flows, like Cæsar's Arar, *incredibili lenitate*, nevertheless it *does* flow, as a few months' inspection would shew.

Yet another way in which we can slow down the process. I have here a circular frame some twenty inches in diameter across which has been stretched, tambourine fashion, a sheet of india-rubber, held in position by a serving of cord and a tourniquet. This is suspended from a convenient bracket, and you see that if water is slowly run into the tambourine, the rubber first assumes a lenticular form and then at one stage, when the gigantic drop is holding several gallons of water, it takes on a position of equilibrium in which a

pronounced waist is seen in the profile of the drop.*
(Plate I, A.) This position, impossible as an equi-
librium position in a water drop, is possible here
because the tension in a rubber surface increases
with the extension, whereas the tension in a liquid
surface is, under conditions of constant tempera-
ture, independent of the extent of the surface.

Even in 1886, it was possible, as this photograph
shews (Plate I, B.), to take an instantaneous photo-
graph of the detachment of a drop of ink from a
funnel.† To-day the progress of high speed kine-
matography has made it possible to follow the
details of the process with the camera. A high-
speed camera developed in the Bell Telephone
Laboratories in conjunction with the Eastman
Kodak Company, makes it possible to take photo-
graphs at a normal rate of a thousand per second
and, by overvolting the motor, we have been able
to increase this rate to nearly two thousand per
second. Obviously at these speeds the intermittent
jerking of the film through the camera is impossi-
ble, and the film hurtles past the lens at a uniform
speed of close on thirty miles an hour when the
motor is overvolted. Between the lens system
and the vertically descending film is a prism which
can be rotated about a horizontal axis at a
maximum speed of eighty thousand revolutions
per minute, and this prism permits twice in each
revolution of the passage of light from lens to

* This experiment was shewn by Lord Kelvin in the course of
his Friday evening discourse of January 29th, 1886. (Proc. Roy.
Inst. Vol. XI. p. 499). If his biographer (*Life of Lord Kelvin*,
p. 854. Vol. II) is a veracious chronicler the growth of the drop
" furnished an exciting episode in the lecture, which culminated
when finally the elastic film gave way and the drop burst over the
lecture table, splashing the nearer members of the fashionably
attired audience."

† Reproduced from the account of Kelvin's lecture given in
" Popular Lectures and Addresses." Vol. I, p. 47.
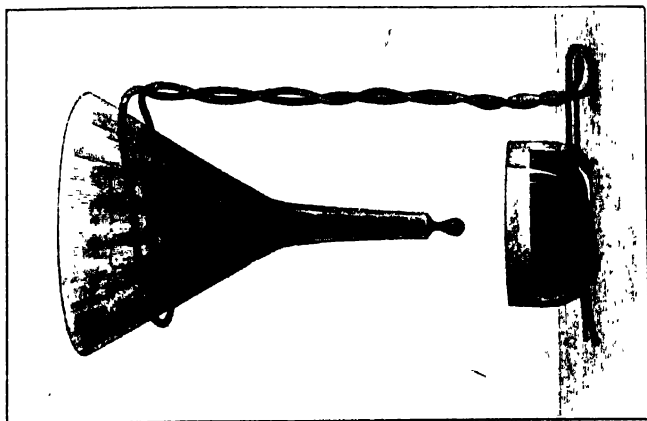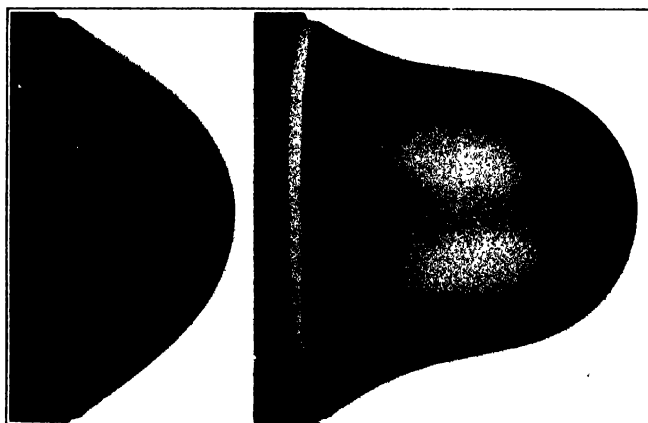
PLATE I.

film, and therefore exposes an image on the film for a period of the order of one five-thousandth of a second.   There is no special difficulty in the illumination of the object ;  in all the experiments to be described a thousand watt lamp and a simple projector gave all the illumination necessary.   A special feature of the camera is a two-dialled clock ;  one of the dials rotates once a minute and registers seconds ;  the other dial rotates once in a second and is graduated into five hundred divisions in such a way that thousandths of a second may be estimated.   An auxiliary prism throws the images of these dials on the film, and the time relations of the phenomenon under observation may therefore be registered with very high accuracy.

Readers of the fantasies of Mr. H. G. Wells will remember that delightful story in which the hero takes a drug which temporarily alters his time scale of living.   Under its influence he writes a three-hour article at what he presumes is his normal rate ;  when the effect of the drug has worn off he finds that he has accomplished his task in a few minutes.   He takes a walk, and notices a bee flapping its wings lazily in the wind ;  he accelerates his speed, and is brought up by a smell of burning—the rapidity of his motion through the air has caused his trousers to singe.

Suppose, then, we make an attempt to enter a world whose time scale is such that we can study the manner in which a drop detaches itself from a tap—we can do this by running a film taken by the high-speed camera through the projector at a rate of fifteen or twenty pictures a second.   We have altered our time scale in the ratio of about a hundred to one, as can well be seen in the films of a stop-watch photographed at a rate of sixteen

pictures a second and then photographed by means of the high-speed camera. The two films are run through at the standard rate of sixteen a second and you see that in the high-speed film it takes much more than a minute for the hand of the stop-watch to cover the distance which is marked as one second.

Let us follow in this way (Plate II, B.) the circumstances of the detachment of a drop, and you see that all the stages—the lenticular form, the waist formation, the drawing-out of a long neck and the final detachment of the drop and its accompanying satellite—can be followed with the greatest ease.

But the most interesting application of kinematography of this type which I have as yet been able to make, is the study of the beautiful phenomena described some thirty or forty years ago by Prof. A. M. Worthington, who photographed the splash of a drop of liquid into liquid, of a solid sphere into a liquid, and of a liquid drop falling on to a solid plate.* That something odd happens when, *e.g.* a drop of water falls on a smoked glass plate is well seen on these slides, the radial striations on which shew that the drop has been throwing out arms in what, at first sight, seems a very queer fashion.

Worthington investigated the phenomenon by taking an instantaneous photograph of, say, a drop of water falling into milk at a certain stage of its fall and then, by an ingenious timing device, ensured that a second drop, released under conditions identical with those of the first drop, should be illuminated instantaneously at a stage of its fall one five-hundredth of a second later

* Worthington, "A study of splashes," 1908.

A  B  C  D  E  F
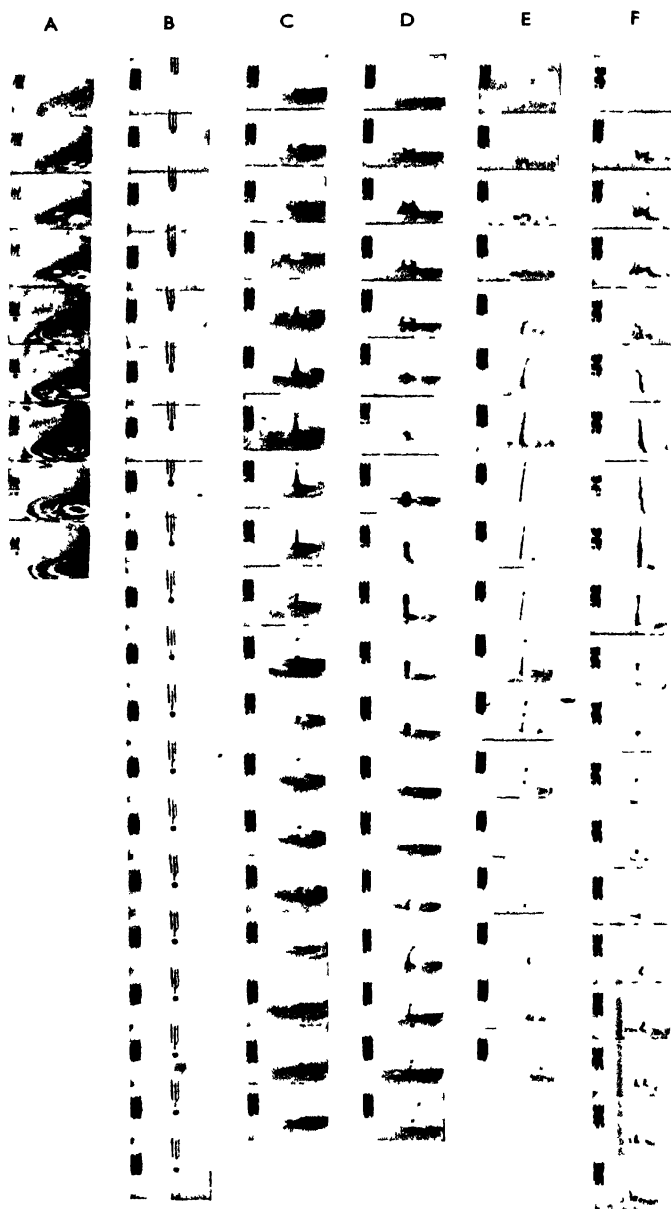
PLATE II.

than the stage at which its predecessor was illuminated. And so on. Worthington, in fact, took photographs of different drops at varying stages of their careers, and from the photographs pieced together the life history of a single drop. The high-speed camera enables us to do this directly, and frees us from certain obvious complexities introduced by the varying behaviour of different drops—complexities which, on occasion, proved very puzzling to Worthington.

The film (Plate II, A.) which shews the splash of a water drop into water is not perhaps so spectacular as some of the others but it has a mild historical interest inasmuch as it is, I think, the first film to shew the phenomenon, and it was actually our first film. taken with no special precautions as regards stopping of the lens and illumination. Addition of a little milk to the water brings out the detail, and it is most interesting to note how closely the story follows that unfolded by Worthington's pictures. We see the effect of a low fall, about 40 centimetres (Plate II, C.) and of a high fall, about a metre (Plate II, D.), the high fall shewing very beautifully the stages in the process of bubble formation on the surface of the liquid. The next films (Plate II, E. and F.) shew the effect of a low and of a high fall due to an ordinary rough marble sphere falling into the liquid. The whole story of these events is told in something of the order of half a second, and the slow motion projection enables us to multiply this period by a factor of about a hundred.

The splash of a drop of mercury on a glass plate is a little more difficult to follow—it is over in about a twentieth of a second and, as you see, even the high-speed camera is not quick enough

to enable us to grasp the complete detail. Nevertheless, we hope to be able to speed up the camera still further, and this gain of speed, with the aid of the technique of the photographer, may make it possible to shew to an audience even so evanescent a phenomenon as this.

[A. F.]

# Friday, March 23, 1934

SIR FREDERICK BERRYMAN, D.C., J.P.
*Manager and Vice-President, in the Chair*

THE LORD RUTHERFORD OF NELSON,
O.M., F.R.S., M.R.I.
*Professor of Natural Philosophy*

## THE NEW HYDROGEN

FOR more than a century scientific men believed with confidence that pure water was a well defined chemical substance, $H_2O$, of molecular weight 18. This belief was shown by the fact that the unit of mass, the kilogramme, consisting of a cylinder of platinum-iridium, was initially chosen to be of the same mass as 1,000 cubic centimetres of water at the temperature of maximum density. Subsequent measurements showed that this was slightly in error, so that the unit of mass was defined in terms of the metal standard. It was only about four years ago that this confidence was disturbed as a result of the study of the isotopic constitution of oxygen. Instead of being a simple element of mass 16, oxygen was found to contain in small quantity isotopes of masses 17 and 18. It was clear from this that pure water must contain some molecules of weight 19 and 20 as well as the normal 18. Since, however, it seemed very unlikely that the proportion of the isotopes could be sensibly changed in the processes of preparation of pure water, this result, while of much theoretical interest, did not appear to have any practical importance.

As a result of investigations during the last two years, there has been a revolutionary change in our ideas of the constancy of the constitution of water. This has resulted from the discovery that a hydrogen isotope of twice the normal mass is always present in preparations of ordinary hydrogen. While this isotope of mass 2 exists only in small proportion, about 1 in 6000 of the main isotope of mass 1, yet, on account of the marked difference in mass of the two components, the relative concentration of the two isotopes can be varied in a marked way by various physical and chemical processes. This is seen by the fact that we are now able to obtain preparations of water in which the isotope of hydrogen of mass 1 is completely replaced by the isotope of mass 2. The density of this heavy water is about 10 per cent greater than ordinary water ; while its freezing point is 3·8°C, and its boiling point 1·42°C higher. While in outward appearance this heavy water resembles ordinary water, yet in general its physical and chemical properties show marked differences. Not only does the vapour pressure vary markedly from the normal but the latent heat is considerably higher. Both the surface tension and specific inductive capacity are lower, while the viscosity is much greater.

It is of interest to indicate briefly the almost romantic history of this rapid advance in knowledge, and to note that there are certain points of analogy between the discovery of heavy hydrogen and the discovery of argon in the atmosphere by the late Lord Rayleigh. In both cases the clue to the discovery depended on the recognition of the importance of small differences observed in accurate measurements of density.

When the relative abundance of the isotopes of oxygen was first measured, Birge and Mendel showed that there was a slight discrepancy—only about 1 in 5000—between the ratio of the masses of the atoms of hydrogen and oxygen, measured by Aston using the method of positive rays, and the ratio deduced by direct chemical methods. They concluded that this small difference was greater than the probable experimental error in the measurements, and in explanation suggested that hydrogen might contain in small quantity—about 1 in 4000—an isotope of mass 2. Let us consider for a moment how the presence of such an isotope could be demonstrated by direct experiment. Both the $H^1$ and $H^2$ isotopes would have the same nuclear charge of 1, and have one external electron, and would thus be expected to give the same type of optical spectrum under the influence of the electric discharge. It is to be remembered, however, that the electron, whose movements when disturbed give rise to its characteristic radiations, is coupled to the nucleus and that the rates of vibration, although mainly governed by the nuclear charge, are slightly affected by the mass of the nucleus itself. On account of the greater mass of the $H^2$ isotope, it can be readily calculated that the Balmer lines in the spectrum of heavy hydrogen should appear slightly displaced towards the red. In the case of the $\alpha$ line, the displacement amounts to 1·78 Angström units. When an electric discharge is passed through ordinary hydrogen, weak satellites should thus appear on the side towards the red. The presence of such weak satellites in the right position was first detected in experiments made for the purpose by Urey, Brickwedde and Murphy. The intensity

of the satellite compared with the strong $H_\alpha$ · line was difficult to measure with certainty but was found to be of the order of 1 to 5000. Experiments were then made to enrich the $H^2$ isotope by fractional distillation of liquid hydrogen and with some success. Another important observation was made by Urey and Washbourn, who found that the water in old electrolytic cells contained a larger proportion of heavy hydrogen than the normal. The concentration of $H^2$ was found to be rapidly enriched by continued electrolysis. This gave the key to a successful method of obtaining heavy hydrogen in quantity. The processes involved were carefully investigated by Lewis and Macdonald, who carried out the electrolysis of water on a comparatively large scale. Nickel electrodes were used, with sodium hydroxide as an electrolyte. In general, it was found that the escape of $H^1$ during electrolysis was from 5 to 6 times faster than that of $H^2$ relative to their concentrations in the solution. There was in consequence a steady accumulation of the heavy isotope in the water in the process until nearly pure heavy water was obtained. Assuming that the initial concentration of $H^2$ in the water was 1 in 6000, about 1 cc. of pure heavy water should be obtained by electrolysis of 6 litres of water.

Lewis succeeded in preparing many cubic centimetres of heavy water in which ordinary hydrogen was present in very small quantity, and with his collaborators investigated the main physical and chemical differences between heavy water and ordinary water, to some of which I have already referred. Our congratulations are due to our American colleagues for the masterly way they have opened up and developed so rapidly this

new field of knowledge, which will certainly prove of great scientific and practical importance in many directions in the near future. Professor G. N. Lewis of the University of California, who was the first to prepare nearly pure heavy water, generously presented samples of this water to a number of investigators, not only in his own country but in Europe, in order to give them an early opportunity of testing its properties. I am personally much indebted to him for a sample of this heavy water with which we were able to make a number of experiments on the transformation of matter to which I shall refer later.

We are all aware of the important part that hydrogen plays in many chemical compounds and particularly in organic molecules. When reasonable supplies of heavy water are available to the experimenter, there will no doubt be great activity in preparing and studying many compounds in which $H^1$ in the molecule is wholly or partly replaced by $H^2$. Already a few investigations have been carried out, for example, with ammonia and with hydrogen iodide in which $H^1$ is replaced by the heavy isotope. It has been found that in mixtures of light and heavy hydrogen gas, the atoms interchange on a nickel surface at a temperature of about 600°C, and the conditions of equilibrium and heat evolution have been investigated. During the next few years we may expect an intensive study to be made of the change of properties of compounds in which heavy hydrogen is used. It will be of particular interest to examine the changes in the rates of reaction at different temperatures when heavy hydrogen is substituted for ordinary hydrogen.

The discovery of the new water will be of great importance in another direction, viz. its effect on the processes occurring in animal and plant life. There has not yet been sufficient time to make more than a few preliminary experiments in this field, and then only on a small scale. Lewis finds that seeds of a certain tobacco plant did not germinate in pure heavy water but did so when the concentration of heavy hydrogen was about one-half. In experiments by other observers, well-defined physiological effects have been obtained for quite small concentrations of heavy hydrogen in water. Further observations in this highly important field of enquiry will be awaited with much interest.

It is generally recognised that the new hydrogen will prove of so much general importance to chemistry and physics that it is desirable to give it a definite name and symbol. Professor Urey, its discoverer, has suggested that the isotope of mass 1 should be called " protium " and the isotope of mass 2 " deuterium " ; while the nucleus of heavy hydrogen, which has already been found very efficient as a projectile in transforming matter, should be called " deuteron " or " deuton." The question of a suitable nomenclature is one of general importance to scientific men and deserves careful consideration. The name diplogen (διπλοῦς, double) for $H^2$ and diplon for the nucleus seemed to find some favour in England as an alternative. The symbol D for the heavy isotope seems appropriate.

While diplogen (or deuterium) may be separated in quantity from heavy water in nearly a pure state, it is of interest to refer to another method of separation employed by Hertz. By utilising a
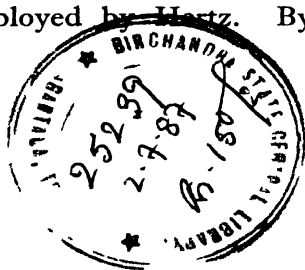
special diffusion method devised by him, he has been able to separate from ordinary hydrogen gas about 1 cc. of diplogen in such purity that the Balmer lines of hydrogen were not visible in its spectrum. With such pure material, it should be possible to study in detail the complicated band spectrum of diplogen and compare it with that of hydrogen.

We have not so far considered the question of the nuclear structure of diplogen and its relation, if any, to that of ordinary hydrogen. We first of all require to know its mass with accuracy ; this has been measured by Bainbridge, using a modification of the positive ray method, who found that the mass of the atom is 2·0136 while the mass of the hydrogen atom is 1·0078 in terms of the mass of the isotope of oxygen taken as 16. This mass is slightly less than the combined mass of two H atoms. Sufficient evidence is not yet available to decide whether the D nucleus is simple or composite, and there are a number of possible combinations to consider between the four units, the electron, positron, neutron and proton. If we assume, as seems not unlikely, that the D nucleus consists of a close combination of a proton with a neutron, it can be shown from the masses concerned that its binding energy should be somewhat less than 1 million volts, if we take the value 1·0067 for the mass of the neutron as estimated by Chadwick. If this be the case, we should expect the diplon to be broken up occasionally into a proton and neutron as a consequence of a close collision with a fast α-particle. Experiments to test this have so far yielded negative results. If this dissociation occurs at all, the probability of such an event must be very small. Lawrence,

from a study of the bombardment of elements by diplons, suggests that the diplon may break up into a proton and neutron in the strong electric field close to the bombarded nucleus, but the interpretation of his results is not yet certain. At the moment, therefore, the experimental evidence is insufficient to give a definite decision with regard to the structure of the diplon.

By comparing the scattering of $a$-particles when passing through diplogen and hydrogen gas, Mr. Kempton and I have found that as the result of a head-on collision with an $a$-particle, the recoiling diplon travels about 8 per cent farther than the proton in a corresponding collision. Such a result is in agreement with calculation. It also seems clear that the field of force round the diplon must be very similar to that of the proton, although it may be expected that some differences would be shown for very fast $a$-particles if the diplon is composite as we have supposed.

*Transmutation of Elements*

The discovery of heavy hydrogen has provided us with a new form of projectile which has proved markedly efficient in disintegrating a number of light elements in novel ways. It was a very fortunate coincidence that when Professor Lewis had prepared some concentrated diplogen, his colleague in the same University, Professor Lawrence, had available his ingenious apparatus for producing high speed protons and other particles with an energy as high as two million volts. When diplogen was substituted for hydrogen, the diplon ($D^+$) was found to be about 10 times as efficient in promoting some transformations in lithium as $H^+$ of equal energy. It will be

remembered that two years ago Cockroft and Walton* found that lithium when bombarded with fast protons was transformed with the emission of swift α-particles. It seems clear that in this case the lithium isotope of mass 7 is involved. A proton is captured by the nucleus and the resulting nucleus breaks up into two α-particles, ejected in nearly opposite directions according to the relation

$$\text{Li}_3^7 + \text{H}_1^1 \rightarrow \text{He}_2^4 + \text{He}_2^4$$
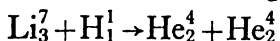
The emission of other particles of short range has also been observed but the exact nature of the transformation which gives rise to them is not yet clear.

When lithium is bombarded with diplons instead of protons, different types of transformation occur. In one case it seems the lithium isotope of mass 6, after capturing a diplon, breaks up into two α-particles according to the equation

$$\text{Li}_3^6 + \text{D}_1^2 \rightarrow \text{He}_2^4 + \text{He}_2^4$$
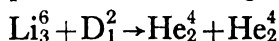
In this case also, as has been shown beautifully by the expansion photographs obtained by Dee and Walton†, the two α-particles are shot out in opposite directions and with a speed greater than the swiftest α-particle from radioactive substances.

Still another interesting type of complex transformation occurs in this element. Oliphant and Rutherford‡ observed that lithium when bombarded by diplons gave, in addition to the group of fast α-particles first observed by Lawrence, a distribution of α-particles of all ranges from 7·8 cms to 1 cm in air. It is believed in this case that the isotope of mass 7 captures a diplon and

* Proc. Roy. Soc. A. Vol. CXXXVII, p. 229, 1932.
† Proc. Roy. Soc. A. Vol. CXLI, p. 733, 1933.
‡ Proc. Roy. Soc. A. Vol. CXLI, p. 722, 1933.

then breaks up into two α-particles and a neutron according to the relation

$$Li_3^7 + D_1^2 \rightarrow He_2^4 + He_2^4 + n_0^1$$

This transformation is in close accord with the conservation of energy when the change of mass and the energies of the expelled particles are taken into account. The emission of neutrons from lithium has been observed by Lauritsen and also in our experiments. In addition, Lawrence has shown that a number of other light elements give rise under bombardment to groups of fast protons and in many cases also to α-particles and neutrons. While the interpretation of the experimental results is as yet only clear in a few cases, there can be no doubt that the use of heavy hydrogen will prove invaluable for extending our knowledge of transformations and thus in helping to throw light on the structure of atomic nuclei.

The importance of this new projectile in studying transformations is well illustrated by some recent experiments made in Cambridge by Oliphant and Harteck.* When diplons were used to bombard compounds like ammonium chloride $NH_4Cl$, and ammonium sulphate $(NH_4)_2SO_4$, in which ordinary hydrogen was in part displaced by diplogen, enormous numbers of fast protons were found to be emitted, even for an accelerating voltage of 100,000 volts. In fact the number of expelled particles is far greater than that observed in any other type of transformation at this voltage. The main groups of expelled protons had a range in air of 14 cms., corresponding to an energy of 3 million volts. In addition to this group, another strong group of singly charged particles

* Nature, Vol. CXXXIII, p. 413, 1934.

of range in air only 1·6 cms. was observed. Both of these groups contain equal numbers of particles. In order to account for these observations, it seems likely that, as the result of a close collision, the diplon occasionally unites with the struck diplon to form a helium nucleus of mass 4 and charge 2, but containing a large excess of energy over the normal helium nucleus. The new nucleus is in consequence explosive and breaks up into two parts, one a fast proton and the other a new isotope of hydrogen $H_1^3$ of mass 3. If this be the case, the proton and $H^3$ nucleus should fly apart in opposite directions. It can be simply calculated that the range of the recoiling $H^3$ nucleus under these conditions should be 1·7 cm.—a range agreeing closely with that actually observed. The changes occurring are illustrated by the equation

$$D_1^2 + D_1^2 \rightarrow He_2^4 \rightarrow H_1^3 + H_1^1$$

From the known masses of D and $H^1$ and the energy of the observed motion of the $H^1$ and $H^3$ particles, it can be deduced that the mass of this new hydrogen isotope is 3·0151.

In these experiments large numbers of neutrons are also emitted. It appears probable that these arise from another mode of disintegration of the newly formed helium nucleus according to the relation

$$D_1^2 + D_1^2 \rightarrow He_2^4 \rightarrow He_2^3 + \text{neutron}$$
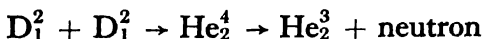
an isotope of helium of mass 3 and a neutron being expelled in opposite directions. There is strong evidence that such an isotope of helium also appears when the lithium atom of mass 6 is bombarded by protons, and from this transformation it appears that the mass of this isotope is 3·0165. It is quite likely that the helium nucleus

of mass 3 formed in this way is unstable and may possibly break up into $H_1^2$ and a positive electron. While the conclusions outlined above are to some extent provisional and require confirmation by other methods, there can be no doubt that the effects which follow the collisions of a swift diplon with another are of much importance and interest in throwing light on possible modes of formation of some of the lighter nuclei.

It is of interest to speculate why the heavy isotope of hydrogen appears in many cases far more effective for equal energies in producing transformations than the lighter isotope. On the general theory of transformation proposed some years ago by Gamow, it is to be anticipated that for equal energies of motion the diplon, on account of its heavier mass, would have a smaller chance of entering a nucleus than the swifter proton. It may be, however, that normally only a small fraction of the protons which actually enter a nucleus is able to cause a veritable transformation, the others escaping unchanged from the nucleus. On this view, the greater efficiency of the diplon in causing transformation .may be due to the fact that a much larger fraction of those which enter the nucleus are retained by it, leading to a violent disintegration of its structure. It may be too that the diplon on entering a nucleus breaks up into its component parts. The appearance of the proton as well as the neutron in some of the transformations may be connected with the composite structure of the diplon.

In this address I have endeavoured to give in a simple way an account of our knowledge of heavy hydrogen which has been gained in the past year and to indicate the great importance of this

discovery to science. This new hydrogen will undoubtedly prove of great value in many ways to Physics and Chemistry and probably also to the Biological Sciences. There are already indications that much interesting information may be obtained by the application of this new substance to the study of processes in animal and plant life.

In the course of the lecture, experiments were shown to illustrate the differences in freezing point and in vapour pressure between ordinary and heavy water, and the differences in heat conductivity between ordinary and heavy hydrogen. For the first time experiments were made to show the artificial transformation of lithium by protons and diplons of energy corresponding to about 100,000 volts. The enormous emission of fast protons when ammonium sulphate containing heavy hydrogen was bombarded by diplons was clearly shown by counting methods. The transformation apparatus was designed and operated by Dr. Oliphant, while Messrs. Watson & Sons loaned an installation to provide a steady potential of 100,000 volts to accelerate the ions.

[R.]

## Friday, April 20, 1934

GEORGE CLARKE SIMPSON, C.B., D.Sc., F.R.S.,
*Manager and Vice-President, in the Chair*

P. M. S. BLACKETT, M.A., F.R.S.
*Professor of Physics in Birkbeck College, University of London*

# COSMIC RADIATION

IN this lecture I hope to do my best to justify the enthusiasm which prompted an American to write the following passage :

"The subject of Cosmic Radiation is unique in modern physics for the minuteness of the phenomena, the delicacy of the observations, the adventurous excursions of the observers, the subtlety of the analysis and the grandeur of the inferences."

The beginning of the story may be put as far back as 1785, when Coulomb demonstrated that the actual air surrounding an electrified body would conduct electricity. The modern development of the subject really started with the work of Elster and Geitel, and C. T. R. Wilson about 1900, who independently showed conclusively that ordinary dry air in a closed vessel was not a perfect insulator for electricity but possessed a small though definite electric conductivity.

It is a remarkable tribute to the brilliant physical intuition of C. T. R. Wilson, that in 1901 he actually wrote :—

" Experiments are now undertaken to test whether the continuous production of ions in

dust-free air could be explained as being due to radiation from sources outside our atmosphere, possibly radiation like Röntgen Rays or like Cathode Rays, but of enormously greater penetrating power."

This indeed was an inspired guess and it has taken thirty years to answer the question he raised.

From such a slight beginning has sprung a vast subject, studied by hundreds of workers all over the world, and one still full of exciting and baffling difficulties. In the 33 years which have elapsed since these first experiments, certainly over 600 papers have appeared dealing with cosmic radiation. They are now appearing at the rate of nearly two a week. In spite of this great amount of work our ignorance is still vast, but what has been discovered is sufficiently exciting. It appears that the earth is being bombarded from all directions by atomic particles, to be precise, positive and negative electrons of extremely high energy—with energies entirely surpassing anything known before in the laboratory or even suspected to exist in nature.

The most energetic atomic particles produced artificially as yet have energies of a few million electron volts. Of the same order are the energies of the particles ejected spontaneously by radium and radio-active substances. But the particles which are streaming continuously out of inter-stellar space and continuously falling on to the earth's atmosphere have energies certainly more than 10,000 million volts—perhaps as high as 100,000 million volts. Where do they come from ? We do not know. How are they produced ? We do not know. It is not that we have

several tenable theories of their origin and that we merely do not know which theory is right. No, it is not that ! We have no tenable theory at all ! We do not know, even, if they could be produced by matter obeying the laws known to us to-day. It is not impossible that the cosmic rays are old archæological remains, dating from a time when the world was young and very different. We cannot yet say. But it is at any rate certain that cosmic rays are an important constituent of the whole universe ; and they may prove to be the key by which the evolution of the universe will become finally understood.

In this lecture I intend to give a general survey of the known facts concerning the cosmic radiation.

The first experiments consisted in measuring the rate of discharge of a gold leaf electroscope under conditions which excluded the possibility of an electric leak taking place across the solid insulator supporting the charged electrode.

If the electroscope is shielded on all sides by many centimetres of a heavy material such as lead, the ionisation falls to about 2J, that is, two ion pairs per c.c. per second. This is interpreted as due to the absorption by the lead of the gamma radiations from the radio-active substances in the ground and in the surrounding buildings. To find out if all the ionisation could be due to this cause, ionisation chambers were taken up in balloons. The results showed that the ionisation due to the penetrating radiation was much greater high up in the atmosphere than on the ground. This important result was established by the work of Gockel, Hess and Kolhörster in the years 1911 to 1913.

Ionisation chambers were also lowered down into deep lakes, and when this was done it was found that the ionisation was very much less than at the surface. Kolhörster carried out experiments in lakes in 1923, which were followed by those of Millikan and Cameron, Steinke and many others. The experiments of Millikan and Cameron were particularly extensive and were carried out in lakes on mountains at various heights above sea level.

In the last few years this type of investigation has been brought to some degree of finality by the beautiful work of Regener. He has lowered a self-recording ionisation chamber 230 metres below the surface of Lake Constance, and has sent another attached to a free balloon up to 25 km. above the ground. At this height the atmospheric pressure is only 22 mm. of mercury, so that the mass of air still above the ionisation chamber only amounted to 3 per cent of that at ground level. The ionisation rises slowly at first from its value of 2J at sea level, then as one ascends it increases faster and then again more slowly, finally tending to a maximum value of about 330J near the top of the atmosphere. In the experiments carried out in Lake Constance the ionisation was found to decrease continuously to the greatest depth reached, which was 230 metres below the surface. At this depth the ionisation was less than 0·01J. This is less than one thirty thousandth of the value found near the top of the atmosphere.

These results made it certain that the ionisation must be due to a stream of rays coming downward through the atmosphere with power to penetrate at least 240 metres of water. These rays must clearly either arise in the very highest parts of the

earth's atmosphere or they must be of extra-terrestrial origin. The study of the variation of the rays with latitude makes it quite certain that the latter hypothesis is the true one.

Another way of investigating these rays is to use a Geiger counter. When two counters are arranged parallel to each other and some little distance apart, it was found by Bothe and Kol-hörster that they discharge simultaneously more often than can be explained by the laws of chance. These simultaneous discharges or coincidences are attributed in general, to the passage of a single fast particle through both counters. If blocks of absorbing material are placed between the two counters the number of coincidences is reduced, but only slightly. Coincidences were obtained by Rossi even when the counters were separated by as much as a metre of lead. A considerable fraction of the ionisation at sea level must, therefore, be due to particles which have a range of over a metre of lead.

The angular distribution of the fast particles producing the coincidences can be obtained by rotating the plane containing the two counters .about a horizontal axis ; the counters themselves are placed parallel and at a fixed distance apart. This experiment has been done by Bernardini and by Johnson. The distribution is found to have a very marked maximum in a vertical direction. This result finds a simple interpretation on the assumption that the flux of radiation inci-dent on the top of the earth's atmosphere is isotropic, and that the observed concentration in a vertical direction is a result of the greater absorption of the radiation which comes inclined through the atmosphere.

Experiments have been carried out to see if the intensity of the rays varies with the time. The results show that the hard components are quite constant. This shows that the rays must be isotropic in space.

The investigation of the variation of the penetrating radiation over the earth's surface is of the first importance for the theory of the origin of the rays. If the radiation is of extra-terrestrial origin, and is electro-magnetic in nature, it is to be expected that the ionisation will be uniform over the earth's surface. While if the radiation consists of charged particles, these will be deflected by the earth's magnetic field in such a way as to produce a greater intensity near the poles than at the equator.

Quite recently the position in regard to the experimental evidence for a variation of the ionisation with geographical position has become much clearer. This has been largely due to the work of Clay, and of Compton and his collaborators. Recently there has been also the interesting work of Auger, Leprince-Ringuet and of Rossi, with Geiger counters instead of ionisation chambers.

The first definite evidence for a variation over the earth's surface was the work of Clay, who measured the variation of intensity of the ionisation during several voyages from Europe to Java, using an ionisation chamber constructed by Steinke. His first results obtained in 1928 have been confirmed by his later and more precise work.

A very elaborate series of experiments in many parts of the world, and at various heights above sea level, was organised by A. H. Compton in

1930 with the assistance of many collaborators. The measurements were made at 69 stations distributed over the earth's surface between latitudes 78 N. and 46 S. The method of measurement was to compare the ionisation due to the penetrating radiation with that due to a standard radium source. The results are of decisive importance.

The main results are as follows : at sea level the intensity is a minimum at the magnetic equator, but rises to a value about 14 per cent higher in geomagnetic latitude of 50 N. and S. and then remains constant towards the magnetic poles. The increase is greater at high altitudes, being about 33 per cent greater at an altitude of 4,300 metres. It is not certain whether the ionisation at high altitudes becomes constant at the same latitudes of 50 N. and S., or whether it continues to increase at still higher latitudes.

These results prove conclusively that a considerable part at least of the radiation producing the ionisation must consist of charged particles.

The experiments already described were carried out with ionisation chambers and with Geiger counters. The first gives the total ionisation, the second the number of particles producing the ionisation. Information of a different kind is provided by the cloud method of C. T. R. Wilson. This gives information as to the actual nature of the ionising particles by condensing water drops on the ions produced by them and so allowing their tracks to be photographed.

In 1929 Skobelzyn first found on cloud photographs the tracks of particles of very much greater energy than any associated with ordinary radioactive substances. Subsequent work with cloud

chambers by Anderson, Kunze, Blackett and Occhialini and others has given considerable information as to the nature of the particles producing the ionisation.

Since the rays are known to be coming mainly downwards, the cloud chamber is generally arranged with its plane vertical, and is placed in a magnetic field in order to bend the paths of charged particles into arcs of circles. From the amount and direction of the curvature of a track and the ionisation density along it, the nature and energy of the particle producing the track can, in general, be inferred.

As was to be expected from the measurements with counters, many more tracks are found with a nearly vertical rather than a horizontal direction. Both Anderson and Kunze have used large magnetic fields, up to 18,000 gauss, and have found that nearly all the tracks are markedly bent. On the assumption that the particles have the same mass as an electron, their energies can be determined. They find that these energies lie mainly between $10^8$ and $10^9$ electron volts, with, however, a very few with energies as high as $10^{10}$ volts.

About half the tracks are curved in one sense and half in the other. On the reasonable assumption that most of the particles are moving downward and not upwards, this implies that about half are negatively and half positively charged. Since the ionisation along the tracks is definitely nearly the same as that to be expected for fast electrons, it is certain that the majority of the negatively curved tracks are actually due to fast electrons.

A study of those tracks which are curved in the sense for a positive particle led to a striking conclusion. A photograph was obtained by

Anderson in which the direction of motion of the particle could be inferred with certainty since it passed through a metal plate in the chamber, and, by losing energy in the plate, showed which way it was going. The direction of curvature indicated a positive charge, but the track was very much too long for it to be due to a proton, which has a mass 1,850 times that of an electron. So Anderson suggested the possible existence of a positively charged particle with a mass comparable with that of an electron rather than with that of a proton.

Definite evidence that many of the positively curved tracks found in association with penetrating radiation are due to such positively charged particles of electronic mass was also obtained by Blackett and Occhialini from cloud photographs taken in a somewhat different way. The usual procedure had been to take photographs at random. On only a small fraction of such photographs is the track of a cosmic ray particle found. So very many photographs must be taken. However, it was found possible to make the high-speed particles take their own photographs— by arranging that their passage through two Geiger counters should operate, by relays, the mechanism that cause the expansion. On 75 per cent of the photographs taken in this way were found the tracks of high-speed particles. Most of the successful photographs show a single nearly straight track passing through both counters, but about 20 per cent show two or more tracks. That such associated tracks were a common feature of the penetrating radiation had been shown by Skobelzyn. What was surprising about the new work was the variety and complexity of these groups of associated tracks ; with the new method

PLATE I.

photographs were taken showing the tracks of
more than 20 particles of high energy. These
appeared to be produced by some kind of nuclear
explosion. Plate I is one of these photographs,
and shows about 25 tracks diverging downward
from the material surrounding the chamber. Those
that are curved to the left are certainly negative
electrons ; those curved to the right must have a
positive charge, but cannot be protons, since
protons of this curvature would have only a
very short range and would ionise very much
more heavily than is observed. So these tracks
must be due to positively charged particles
with a mass very much smaller than that of ·a
proton.

It is reasonable to assume that most of the
tracks found by Anderson and Kunze to be curved
in a positive direction are due to these " positive
electrons," though some of them may perhaps be
protons. On this assumption, the main part of
the ionisation due to the penetrating rays at sea
level is due to about equal numbers of positive
and negative electrons. Anderson's and Kunze's
results show that there is not much difference
between the distribution over different energies of
the two kinds of particles.

The established fact that the ionisation varies
over the surface of the earth in a way closely
correlated with magnetic latitude makes it certain
that some part at any rate of the incident radiation
consists of charged particles which are deflected
by the earth's magnetic field.

The theory of the deflection of such charged
particles is essentially that given by Störmer in
connection with the theory of Aurora Polaris. Here
the main results only will be given. Electrons of

given energy $E$ can only reach the earth's surface at latitudes greater than some critical latitude $\phi$. Table I gives the values of $E$ and $\phi$ as calculated by Heisenberg from Störmer's theory.

TABLE I.

| Geomagnetic latitude $\phi$ | 0 | 10 | 30 | 50 | 70 |
|---|---|---|---|---|---|
| Electron energy—$E \times 10^{-6}$ volts | 10,400 | 9,850 | 6,600 | 2,600 | 200 |

If, therefore, the observed ionisation at the earth's surface is due to electrons, then they must have energies of the order of $10^{10}$ volts. It is satisfactory to note that a few electrons with energy of this order have been observed by Kunze in an expansion chamber.

The most striking feature of the experimental results is the constancy of the intensity near the poles and the sudden drop at latitudes of about 45° N. and S. It has been pointed out by Clay that this is just what is to be expected in view of the effect of the absorption of the rays by the earth's atmosphere.

If we imagine that electrons of all energies are incident on an earth without an atmosphere, then the number of electrons reaching the surface will increase continuously from the magnetic equator to the magnetic poles. At the equator only electrons of energy greater than $10^{10}$ volts can reach the surface. Near the poles, electrons of much lower energy can also do so. Now the effect of the earth's atmosphere is to prevent all electrons whose energy is less than about $4 \times 10^{9}$ volts from reaching the surface. So from latitude 45°, which is the critical latitude for electrons of

this energy, to the poles, there can be no further increase in the number of electrons reaching the earth's surface. The ionisation should, therefore, remain constant over this region, as it is observed to do.

A more detailed discussion of the orbits described by fast charged particles in the magnetic field of the earth, and of the distribution of their incidence on the earth's surface has been recently given by Lemaître and Vallarta.

To explain the penetration of the rays to depths of more than 200 metres of water, it is necessary to suppose that some electrons are incident on the earth with energies of over $10^{11}$ volts.

At ground level the ionisation seems to be mainly due to negative and positive electrons, with possibly also some protons. Whether the incident radiation has this same composition it is as yet impossible to say. Rossi first pointed out that it should be possible to decide the sign of the charge of the incident particles by observing the ratio of the number of particles arriving on the earth's surface from · an easterly and from a westerly direction.

Quite recently Johnson has reported that experiments carried out in latitude 31° N. and at a height of 2,550 metres, show that the intensity of the radiation coming from the west is as much as 15 per cent greater than that from the east. This result must be interpreted as evidence that the majority of the incident particles are positively charged. It is very probable that the majority of the incident particles are positrons.

C. T. R. Wilson has suggested the possibility that the penetrating radiation may have its origin

in the electric fields of thunderstorms, and he has shown that electrons must be produced in these fields with nearly the whole energy corresponding to the potential difference existing in a thunder-cloud. This potential is usually of the order of $10^9$ volts, occasionally it may be nearly $10^{10}$ volts, but certainly not more. This raises an initial difficulty, since the variation of the ionisa-tion over the earth's surface seems to demand energies considerably greater than $10^{10}$ volts, and the penetration of some of the rays to great depths of water seems to demand still greater energy.

But there is another serious objection to the thunderstorm theory of the origin of the main cosmic radiation. For it can easily be seen that at a given magnetic latitude, no particles which reach the earth's surface can have come from the earth's surface, unless their energy is less than the critical values given in the table. Thus, for instance, at latitude 45° no particle which has energy greater than $4 \times 10^9$ volts can have come from the earth. If, therefore, all the cosmic rays were of terrestrial origin, then no particles of energy greater than $4 \times 10^9$ volts could reach the earth's surface in latitudes greater than 45°. That is, at such lati-tudes no particles at all should be able to penetrate the atmosphere. The intensity at sea level should therefore drop to zero at about latitude 45ᵛ.

It is very difficult to conceive of any other terrestrial mechanism for the rays other than thunderstorms. For instance, the high con-ductivity of the upper atmosphere makes it impossible to suppose that a radial electric field of sufficient magnitude could be maintained there. In addition the observed variation of coincidences

with the angle from the zenith gives evidence against any predominantly radial motion of the particles and is in favour of the assumption of an isotropic incident radiation.

If the extra-terrestrial origin of the rays is assumed, then the extreme constancy in time of the hard radiation makes it necessary to assume that the radiation is completely isotropic in inter-stellar space. The sun is therefore clearly excluded as a possible source, and so are the stars of the galactic system, since they are by no means uniformly distributed with respect to the earth. In the former case, one would expect a marked variation of intensity with solar time, and in the latter case with sidereal time. No such variation exists, at any rate for the harder components.

If the galaxies themselves are distributed at random in space they might be cited as a possible origin, or some process involving inter-stellar matter might also give a sufficient degree of isotropism. This has been the view of Millikan.

It is quite certain that the radiation cannot originate from the interior of any star, since the rays would be degraded to much lower energies during the passage outwards into space. It can only arise from the outer parts of stars, from nebulæ or from the diffuse inter-stellar matter.

It has been argued above that since the radiation falling on the earth is isotropic, it must come from sources distributed uniformly with respect to the earth. An alternative view, first suggested by Regener, is that the radiation may have been produced at an earlier stage in the evolution of the universe, and may have since become isotropic while circling the closed universe ever since. Eddington estimates that the average amount of

matter encountered by a ray which traverses the whole universe is equivalent, as regards absorption, to about 7 cm. of water. This is less than 1 per cent of the absorption by the earth's atmosphere.

In addition to the question of where the rays originate, there is the further question of how they originate. Various possible mechanisms have been suggested. The possibility that the rays have been produced by cosmic electric fields is difficult to reconcile with the isotropism of the radiation, unless their birth is placed in the distant past. The hypothesis of the annihilation of matter hardly gives enough energy, for the energy corresponding to the total disappearance of the mass of a proton only gives $10^9$ volts, and it seems unlikely that heavier complex nuclei could disappear so as to transfer all the energy due to their mass to electrons or photons. The hypothesis of the building up of complex nuclei from simpler constituents certainly cannot give enough energy.

It may be said that all the hypotheses that have so far been made to explain the origin of the penetrating radiation are unplausible on one ground or another. The thunderstorm theory is the only one which retains the rays as a terrestrial phenomenon. But if this theory should prove untenable, as seems likely, then the problem of their origin becomes at once intimately bound up with the whole theory of cosmogony. For it has been estimated by Lemaître that the total energy of the cosmic rays, assuming them to be uniformly distributed throughout space, may possibly be as much as one-thousandth of the total energy of all the matter in the universe. The question of the origin of the rays becomes even more mysterious if, as appears to be the case, the majority of the

incident rays are positrons. It would thus appear, too, as probable that the positron, though so rare, because ephemeral on earth, is a very important constituent of the universe as a whole.

[P. M. S. B.]

Friday, November 2, 1934

SIR ROBERT ROBERTSON, K.B.E., M.A., F.R.S.
*Treasurer and Vice-President, in the Chair*

F. W. ASTON, M.A., Sc.D., D.Sc., F.R.S.
*Fellow of Trinity College, Cambridge*

# ELEMENTS AND ISOTOPES

## [ABSTRACT]

THAT a chemical element could consist of isotopes of different atomic mass was first observed by Soddy when working on the products of radio-activity. Proof that this was true of the elements generally could only be obtained by direct atomic analysis. This was achieved by the mass-spectrograph, and with it the search for isotopes has been carried on continuously for the past fifteen years. Wide differences of properties among the elements necessitate very varied methods of obtaining the atomic rays required for the analysis. In some cases the technical difficulties are great and it was only during the last year that satisfactory results were obtained with the rare earth group. Of the common elements all but four, Palladium, Iridium, Platinum and Gold, have now been analysed, and some 247 isotopes identified, a few by less direct optical methods. Elements of odd atomic number appear curiously limited to two isotopes, but even ones can have many more, eleven in the case of tin. By means of modern instruments it is possible to compare the masses of atoms to one part in ten thousand, an accuracy which it is expected to increase in the near future. These isotopic weights are required in order to test theories of

nuclear structure which have recently become of the greatest importance on account of the discovery of transmutation and of the artificial production of radioactive isotopes.

[F. W. A.]

Friday, November 23, 1934

SIR ROBERT ROBERTSON, K.B.E., M.A., F.R.S.
*Treasurer and Vice-President, in the Chair*

M. POLANYI, M.D., Ph.D.
*Professor of Physical Chemistry in the Victoria University
of Manchester*

## HEAVY WATER IN CHEMISTRY

ONE gram is the weight of one cubic centimetre of water at 4° C. One cubic centimetre of heavy water weighs about 10 % more, that is 1·1 grams.

The molecule of heavy water is composed of hydrogen and oxygen in the same proportion as that of ordinary water ; two hydrogen atoms being united with one oxygen atom. Nor is there anything unusual about the oxygen atom in this heavy water molecule. But the hydrogen is different from ordinary hydrogen. Its atomic weight is 2 instead of 1. It is to this new sort of hydrogen that all the heaviness of heavy water is due. It is this heavy hydrogen discovered by Professor Urey in New York, which interests chemists in heavy water.

At first sight, this interest may well seem unjustified. Heavy hydrogen is not a representative of a new class of substances. It is to be considered as an isotope of hydrogen, which is accompanied by it in the same way as almost every element is accompanied by one or more of its isotopes. Lead for example, which is mainly constituted of atoms weighing 208 units, contains beside these atoms of the weight 203, 204, 205, 206, 207, 209 and 210. In chlorine there is, beside

the main part consisting of atoms of weight 35, another kind of atom weighing 37 which forms as much as one third of the element.

## CONFLICTING DEFINITIONS OF ISOTOPY

These were great revelations when more than twenty years ago Soddy discovered the existence of isotopes, and Aston disclosed, by his mass spectrograph, the isotopic composition of the elements. But in later years, discoveries of new isotopes have ceased to arouse general interest and even when more recently, the three basic elements of organic chemistry and of living matter, carbon, oxygen and nitrogen were discovered to contain a fair amount of heavier isotopes, namely, carbon of weight 13 beside that of weight 12, nitrogen of weight 15 beside that of weight 14, oxygen of weight 18 beside that of weight 16, these discoveries did not arouse great enthusiasm among chemists. Indeed, many excellent chemists of my acquaintance have taken no notice of these new isotopes.

Why then, is the new isotope of hydrogen viewed so differently from other isotopes, that some eminent chemists consider its discovery to be possibly the greatest advance in chemistry made in this century ?

The answer is, because it does not behave as an isotope at all. So much so, that Professor Soddy, the discoverer of isotopy, has in contradiction to the general view, actually repudiated it altogether from the rank of true isotopes.

Professor Soddy upholds the original definition of isotopy, according to which two elements should be called isotopes, if they cannot be separated from one another by any chemical means.

By this standard, the two different hydrogens should certainly not be considered as isotopes. Heavy hydrogen is easily separable from ordinary hydrogen.

I can lay upon the table proof of this, in the form of a specimen of 95 % pure heavy water. This is not a natural product, but was manufactured by Imperial Chemical Industries in England from ordinary water which contains only 1/4000 of heavy water. Evidently a very effective separation of the heavy hydrogen from the ordinary one has been carried out in this case. Also there is no doubt that the process used for the separation is a chemical one.

The preparation consists in a process of electrolysis. The first indications of the separability of the two hydrogens by electrolysis was discovered by Washburn and Urey, who found that when water is decomposed by electrolysis, the undecomposed residue has a somewhat greater density than ordinary water. The purification of heavy water on this basis is due to Professor Lewis of California, who has shown that by decomposing very large quantities of water up to a small residue, almost pure heavy water is obtained.

There is plenty of other evidence for chemical differences between ordinary and heavy water. Generally the compounds of heavy hydrogen react more slowly than the corresponding ordinary hydrogen compounds. The greatest difference has been described by Professor Urey in the reaction between water and aluminium carbide, which leads to the formation of methane. Heavy water reacts 20 times slower than ordinary water.

Why, then, if the two hydrogens are so different, do the chemists generally agree to consider them

as isotopes ? The answer is, because the two hydrogens, although chemically different, are true isotopes with regard to the structure of their atoms.

The amplification of the original definition of isotopes implied in this opinion is the natural outcome of the theory of Rutherford and Bohr on atomic structure. We can illustrate this structural view of isotopy by comparing the atomic models
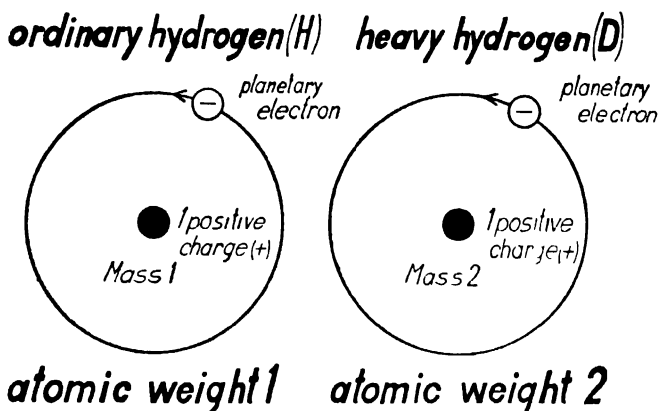


Fig. 1 —The Isotopes of Hydrogen.

of the two sorts of hydrogen atoms. These are shown in Fig. 1 according to Bohr's theory. The two atoms are equal in every respect, apart from the difference in the masses of their nuclei. Two atoms thus related to one another are considered to be isotopes from the structural point of view.

Until the discovery of heavy hydrogen, atoms which have the same structure, and differ from others only in the mass of the nucleus, have always been found to possess identical chemical properties.

It is easy enough to show reasons why this

should be so. The forces originating from an atom are due to the electric field of the charges contained in it. Two atoms with identical electric charges, identically distributed in space, will therefore originate identical forces. So it is to be expected that such a pair of atoms should have equal chemical properties.

The astonishing thing is, that this should not hold for the two sorts of hydrogen atoms, that these two, although originating identical forces, should have different chemical properties. How can the mere difference in nuclear mass cause such marked chemical differences as shown by the two hydrogens ? If mass differences can cause such disparities, why have they never become apparent in other known pairs of isotopes ; why has this mass effect remained undisclosed up to the discovery of heavy hydrogen ?

Only when we can fully answer these questions, shall we be quite justified in considering heavy hydrogen as an isotope of ordinary hydrogen.

Now there is a certain difference in chemical properties caused by mass-differences, which has been well known for a long time, and which we must expect to find more accentuated between the two sorts of hydrogen atoms than between any other pair of isotopes known hitherto. This difference has its origin in the motion in which all particles around us are kept by the *heat* contained in matter. The thermal velocity of a lighter particle is greater than that of a heavier one. A particle moving faster will reach a molecule with which it might react faster than its slower competitor ; it will therefore be found to react quicker, just as light hydrogen reacts quicker than heavy hydrogen.

The chemical differences which arise from thermal velocities will depend on the *ratio* of the atomic masses. This ratio is certainly more marked in the case of the two hydrogens than of any other element. It is 1 : 2 for the hydrogens, while in the case coming next to these, *i.e.*, the pair of lithium atoms Li 6 and Li 7, there is a ratio of only 1 : 1·2. Thus there is a good *prima facie* case for attributing the chemical differences between the hydrogen isotopes to their different thermal velocities.

However this explanation, although it looks so promising at first sight, turns out on closer examination to be an incorrect one.

Firstly : a calculation shows that the differences in thermal velocities are quite insufficient to account for the differences which have been actually found between the reaction velocities of the two hydrogen isotopes.

Secondly : there are some dissimilarities to be described presently between the compounds of the two hydrogens, which prove that these compounds differ in their *energy content*. Consideration of thermal velocities cannot account for such energy differences.

We must, therefore, postulate a cause apart from the differences in thermal velocities, for the explanation of the actual dissimilarities of the two sorts of hydrogen. We will see that this cause is to be found by applying to our problem one of the more recently discovered principles of nature, namely the uncertainty relation of Heisenberg.

### THE LAW OF UNCERTAINTY

The uncertainty principle states, that no information can be obtained about the velocity of a

particle, the position of which is known with absolute accuracy. *Some* information about the velocity might be arrived at, if we admit *some inaccuracy* of position. And thus, the two inaccuracies remain tragically linked together in the formula :

*inaccuracy of position* × *inaccuracy of velocity*
= *constant.*

Our information on position and velocity, have between them a compound inaccuracy which is irreducible.

I will show that from this uncertainty, we can derive a dynamical principle latent in all matter, which acts against a force holding a particle, and in doing so modifies the effects of the force. It will also appear that the effect of this dynamical principle depends on the mass of the particle, and is, therefore, different for two atoms originating identical forces, but differing in mass. We will then see that this is the true reason why the two hydrogen atoms are so different.

An imaginary experiment will lead us to recognise the dynamical principle in question. Suppose we attempt to defeat the uncertainty principle by sheer force. We take an atom and hold it at rest in some fixed position. If we succeed in doing this, we would obviously overthrow the law of uncertainty. The position of our atom would be exactly known and, since we suppose it to be held at rest, its velocity would also be known to be exactly equal to zero.

The law of uncertainty predicts that our experiment will fail. Any force trying to keep an atom in a fixed position would be defeated by a

power given to the particle to defend its uncertainty. It will defend it by starting to vibrate. The tighter we try to hold the atom to stop this vibration, the more violent would the vibration become. No force would be strong enough to keep the particle in place, motionless.

The uncertainty law thus leads to the following postulate : any particle restricted to a definite range of positions is necessarily in motion ; the range of velocities contained in this motion will be the wider the narrower the restriction of positions. Or written as a formula :

*range of position × range of velocities = constant.*

In nature, atoms are restricted in their position when linked up to chemical compounds. Such restrictions, we must conclude, will give rise to an uncertainty motion of the atoms. All molecules will hence contain a certain amount of uncertainty motion, and also, since this motion has kinetic energy attached to it, a certain amount of corresponding energy. We might also postulate, that the more restricted the positions of the atoms in the molecule, are, that is, the stronger the bonds are that hold the atoms in position, the more violent will be the uncertainty motion, and hence the greater will be the energy content of the molecule, due to uncertainty.

Next to bond strength, atomic mass will influence the uncertainty motion. This influence of mass is contained in the constant of the uncertainty formula which can be written :

$$\frac{range\ of\ positions \times range\ of\ velocities}{mass} = universal\ constant$$

Thus the compound uncertainty on the left hand side of the equation is the smaller the larger the mass of the particle. Hydrogen of mass 1 will have a compound uncertainty twice as large as hydrogen of mass 2. Under equal conditions, therefore, the uncertainty motion and the energy of this motion will be larger



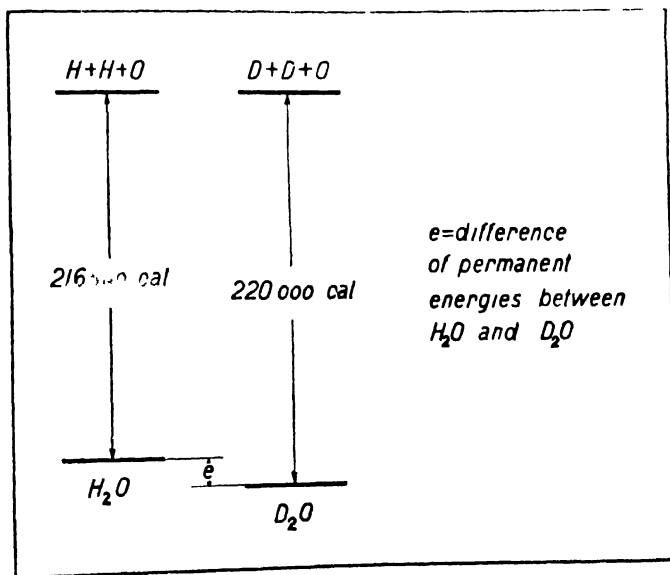FIG. 2.

for light hydrogen than for heavy hydrogen; in corresponding molecules, containing the two sorts of hydrogen, there will be more "uncertainty energy" present when the molecule contains the ordinary hydrogen, than when it contains the heavy isotope.

Compare, for example, ordinary water with heavy water. For ordinary water the "uncertainty energy" amounts to 13097 cal., for heavy water it

is only 9527 cal. Since the uncertainty energy is
only present in molecules, and vanishes when the
atoms are set free, it follows that less work is
needed to break up an ordinary water molecule
into free atoms, than to separate the atoms of heavy
water. This is illustrated in Fig. 2.

From such differences in the energy contents
of the corresponding molecules all the differences
in the chemical properties of the two hydrogens
arise.

I will show this in the remaining part of my
lecture, but before turning to this, I wish to
emphasise two points.

Firstly, that the opinion, attributing the excep-
tional dissimilarity of the two hydrogen isotopes
to the exceptionally high ratio of their masses is
not correct. Suppose a lead isotope should be
discovered having double the mass of ordinary
lead. Such an isotope would be chemically
indistinguishable from ordinary lead, because the
" uncertainty " attached to a particle of the mass
of a lead atom is imperceptible and hence no
variation of this uncertainty can be detected.

My second point is this : the permanent
character of the atomic motion, which is required
to keep up the uncertainty of velocities, should be
clearly perceived. Atoms and molecules are
ordinarily kept in, what may seem to us, perpetual
motion by heat. But heat can be passed on to a
cooler body, or be lost altogether by radiation.
In the distant future, all heat may become lost
by radiation, and all thermal motion may die out.
But beyond that death, the uncertainty motion
will persist for ever.

No atom bound in a molecule can ever find rest
from this motion, nor lose the energy arising
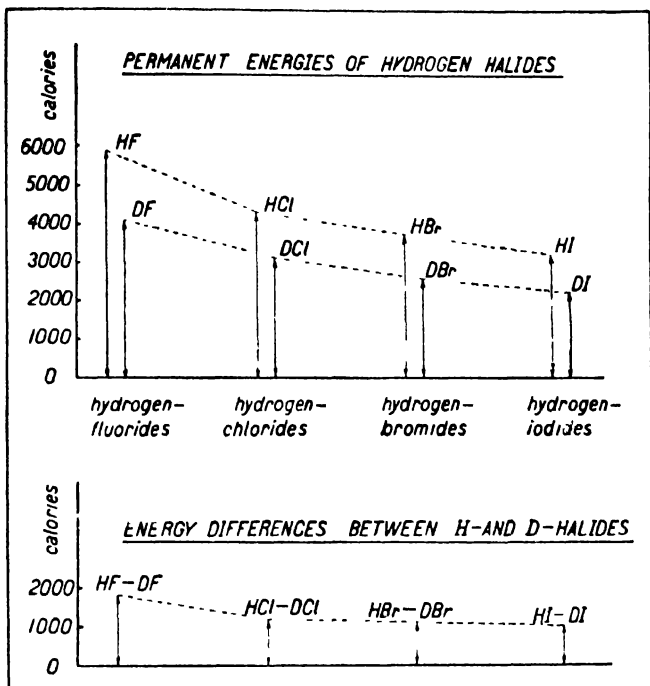
FIG. 3.

from it. We might well call this the *permanent motion*, and the energy corresponding to it, the *permanent energy* of the molecule.

PERMANENT ENERGY AND CHEMICAL
PROPERTIES OF THE HYDROGEN ISOTOPES

We have now to show in what way the differences in permanent energies cause the dissimilarities in the properties of the ordinary and the heavy hydrogen.

The curves in Fig. 3 give a representation of the permanent energies of both the ordinary and the heavy hydrogen halides.

Since the bond strength of the hydrogen halide molecules decreases in the sequence F → Cl → Br → I, we might expect, remembering that the permanent energy is the greater the tighter the bond which holds the atoms in position, that the permanent energy will decrease in the sequence of falling bond strength.

This is well borne out by both curves which show consistently a decrease in the sequence HF ‣ HCl → HBr ‣ HI, and likewise in the corresponding sequence DF ‣ DCl ‣ DBr ‣ DI.

The reduction of permanent energy which has been deduced from the uncertainty principle for the case of H being replaced by D, is also clearly shown.    The D-curve lies everywhere below the H-curve.

The relative depression of the permanent energy is very nearly equal for all four compounds. Consequently, the absolute value of the difference in permanent energies is the greater the higher the permanent energy of the original compound. This relation when connected with the afore-mentioned rule governing the sequence of the permanent energies, leads to the important conclusion, illustrated by the lower part of the figure, that the differences in the permanent energies of corresponding H and D compounds, fall off in the sequence of decreasing bond strength. Or putting it in a more general way, the contrast between two corresponding ordinary and heavy hydrogen compounds will differ from compound to compound, and will be the more marked the firmer the bond by which the hydrogen is linked in the compound.

The energies in Fig. 3 are not measured data, but values calculated from the theory of molecules.

A little further discussion leads us to a very sensitive method of checking these theoretical results.

The diagram shows us that if we replace H by D in HF, the energy will fall by about 1,600 cal., i.e. this amount of energy will be gained. Similarly, if we replace H by D in HI, we gain about 900 cal. We can express this in the chemical equations :

(1)   $D + HF - DF + II + 1600$ cal.
(2)   $D + HI = DI + H + 900$ cal.

By deducting equation (2) from equation (1) we obtain, after a slight rearrangement :

$$HF + DI = HI + DF + 700 \text{ cal.}$$
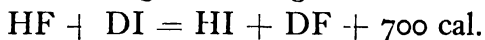
Hence an interchange of H and D between HF and DI is a reaction in which energy is set free. And since reactions always tend to go in the direction in which they produce energy, we might expect that in a mixture of hydrogen fluoride and hydrogen iodide, which have between them a certain amount of heavy hydrogen, the heavy hydrogen will have the tendency to unite with F rather than with I.

An experiment to test this conclusion, could be carried out in the following way. A quantity of heavy hydrogen could be prepared by decomposing heavy water by electrolysis. We might use one gram of water, containing 1 % of pure heavy water, and by completely decomposing it, produce about one litre of hydrogen, containing 1 % heavy hydrogen. From one half of this we could make with fluorine, one litre of hydrogen-fluoride containing 1 % of DF. The other half would go to form 1 litre of hydrogen iodide containing 1 % DI. We could now let the two gases mix together in a 2 litre vessel, and add a trace of water to catalyse the interchange of

hydrogen atoms between the two gases. On separating the gases and estimating how much heavy hydrogen is contained in each of them, we should find that D accumulates in the hydrogen fluoride, which will contain about $1 \cdot 3$ % of D as against $0 \cdot 7$ % of D in the hydrogen iodide.

By carrying out the experiment at a low temperature, for example $- 150°$ (supposing that an efficient catalyst could be found) the distribution of D would become even more unequal, namely $1 \cdot 8$ % D in hydrogen fluoride as against $0 \cdot 2$ % D in the hydrogen iodide.

### INTERCHANGE REACTIONS
### OF HYDROGEN ATOMS

Such interchanges of H and D between two hydrogen compounds have been the object of numerous studies, especially at the Universities of Manchester and Cambridge.*  Indeed the principal part played by heavy hydrogen in chemistry is in some way or another connected with such interchange processes.

Suppose that we bring together the two gases, hydrogen and hydrogen iodide, and add to these the three liquids, water, benzene and ethyl alcohol, and suppose also that we have appropriate catalysts present to bring about the interchange of the hydrogen atoms between all these compounds, then, after separating the substances, we shall find each of them contains a certain part of the heavy hydrogen present in the mixture.  This characteristic quota of each compound, will specify the relative preference which it gives to D over H.

* In Manchester the work was mainly done by Dr. J. Horiuti ; the work in Cambridge is due to Dr. A. Farkas, Dr. L. Farkas and Professor Rideal.

A list of these quota figures for the five afore-mentioned compounds, is given in the Table in which the units are, of course, arbitrary. From what has been said before, we know that these figures depend on the differences of permanent energy between the ordinary and the heavy compounds. We obtain from them a rather

DISTRIBUTION OF D BETWEEN
DIFFERENT HYDROGEN COMPOUNDS

| hydrogen compound* | specific quota | Reference |
|---|---|---|
| HI | 0·17 | Calculated from known equilibria |
| $H_2$ | 0·33 | A. and L. Farkas Trans. Far. Soc. *xxx*. 1071. 1934 |
| $H_2O$ | 1·00 | (arbitrary unit) |
| $C_6H_6$ | 0·95 | J. Horiuti and M. Polanyi Nature, *cxxxiv*. 377. 1934 |
| $C_2H_5OH$ (hydroxyl group only) | 1·5 | C. E. H. Bawn (unpublished) |

* The symbol H as used here includes both kinds of hydrogen.

intimate knowledge about the permanent energy of different compounds which otherwise would not be easily accessible to measurement.

The capacity of some substances to accumulate a comparatively high quota of the heavy hydrogen present in a mixture, can be utilised in the following way. Suppose we bring hydrogen iodide containing some D into contact with alcohol, then

we will find on separating the two substances, a concentration of D about 10 times greater in the alcohol than in the hydrogen iodide. If we carry out the process at low temperatures, for example at – 80° C., the ratio of the two concentrations will be as high as 30 to 1.

Processes of this kind may promise to be of use for the manufacture of heavy hydrogen. Ordinary hydrogen contains, as I have said, about 1/4000 of heavy hydrogen. To concentrate it from this dilution at a reasonable cost is as yet an unsolved problem.

Suppose we convert the hydrogen gained by decomposing ordinary water into hydrogen iodide, and then pass this hydrogen iodide through alcohol at – 80° C., we would get an alcohol containing almost 1 % of D in its hydroxylic hydrogen. By decomposing the hydroxyl group of the alcohol, e.g. by metallic sodium, a hydrogen containing almost 1 % D would be made free, and it would be easy to arrive at highly concentrated D, by repeating the process once or twice.

In practice this process would probably fail on account of the unavoidable losses of iodine and of alcohol, which would make it fairly expensive.

Similar processes, based on the unequal distribution of D between different substances, will be probably found practicable sooner or later, and may then bring down the price of heavy hydrogen to the point that it could be used in the manufacturing of the more valuable chemical products, such as drugs and dyestuffs.

Another interest attached to the interchange of hydrogen atoms between different hydrogen compounds, lies in the possibility which they afford for

the preparation of the more complicated compounds of heavy hydrogen. It is, of course, not impossible to build up all sorts of heavy hydrogen compounds by synthesizing them from the elements, using heavy hydrogen instead of the ordinary one. But this procedure might prove rather awkward with many very common substances which are not usually prepared by a synthetic process, such as benzene, naphthalene, anthracene. However, it seems easy to prepare the heavy hydrogen compounds corresponding to benzene, naphthalene, etc., by taking the ordinary substances and replacing in them the H atoms by D atoms.

Suppose we want to make benzene with the hydrogen atoms substituted by D atoms, i.e., $C_6D_6$. A synthesis could be carried out by polymerizing synthetic heavy acetylene. It seems much simpler to bring the benzene in contact with pure heavy water, adding an appropriate catalyst to let the two substances exchange their hydrogen atoms. If we take 10 grams of heavy water and one gram of benzene, about 90 % of the hydrogen in the benzene should be replaced in one process. A repetition of this procedure should give us benzene containing 99 % D in its hydrogen. This process is now being tested in Manchester.

The interchange of hydrogen atoms of different compounds has also an interest as a new type of chemical reaction, often related in an interesting way to other " true " chemical reactions.

Take for example, the replacement of H by D in benzene. The quickest way to obtain this replacement is by contacting heavy hydrogen with benzene at room temperature in the presence of a nickel or a platinum catalyst. These catalysts are

well known for their capacity to cause the addition
of hydrogen to unsaturated compounds.   In their
presence ethylene reacts very rapidly with hydro-
gen to form ethane.   Benzene likewise adds on
hydrogen, forming hydro-benzene, although much
more slowly.   The replacement reaction will,
therefore, be accompanied by a hydrogenation
of benzene.   But experience has shown that
the hydrogenation is very much slower than the
replacement.   Only one in a hundred molecules,
reacting in the sense of replacement, reacts also
in the sense of hydrogenation.

The replacement of H by D in benzene, can
also be carried out by contacting heavy water
with benzene.   This reaction proceeds also in the
presence of platinum and nickel catalysts, but it
goes much slower than the interchange between
hydrogen and benzene.   Higher temperatures and
longer times are required when heavy water is
used for replacement ;   there is, of course, no
hydrogenation whatever.

We note that both the hydrogenation and the
replacement of hydrogen atoms represent a trans-
fer of hydrogen atoms to the benzene.   The two
processes differ only in the result obtained by the
transfer of the hydrogen atom.   While in the case
of the *hydrogenation* the transfer results in the
formation of hydrogenated products, like ethane
from ethylene, or hydro-benzene from benzene,
the replacement proceeds without any accom-
panying chemical change.

These alternative reactions can be explained by
the reaction scheme shown in Fig. 4.

A hydrogen atom coming in contact with an
unsaturated molecule, first forms a half hydro-
genated, product.   This substance then, if left to

itself, decomposes by dropping one of its redundant
hydrogen atoms (see upper arrow), whereby there
is at least an even chance that the hydrogen atom
lost is not the same one as had been added before,
and that in consequence, the result is the replace-
ment of a hydrogen atom. This decomposition
of the half-hydrogenated state can, however, be
forestalled if a second hydrogen atom comes up



FIG. 4.

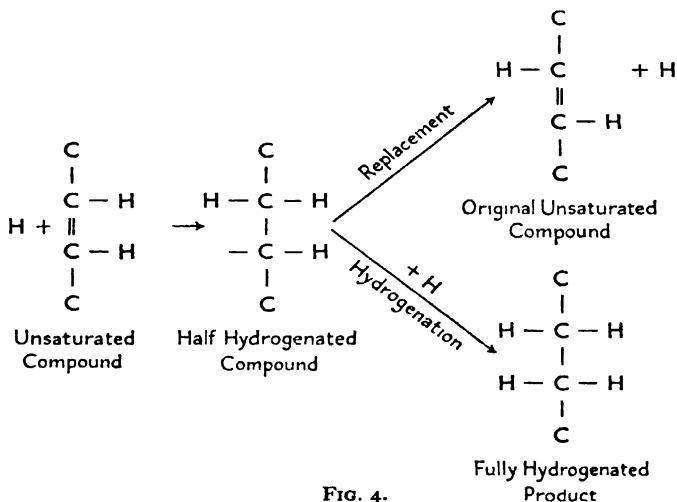before it is accomplished (see lower arrow), and
links up to the half-hydrogeneated molecule,
forming a fully hydrogenated compound.

If this explanation is correct, hydrogenation
will be rare, when the interval between the
approach of the first and second hydrogen atom is
long. In such cases the reaction will result
almost exclusively in replacement of hydrogen
atoms, unaccompanied by hydrogenation.

This conclusion is well borne out by our experiments, which show that while a more energetic action of hydrogen on benzene (when gaseous hydrogen is contacted with it) causes a replacement of hydrogen atoms, which is accompanied by a quite appreciable amount of hydrogenation, no hydrogenation is found when the action of hydrogen is slow, as for instance, when water is the source of the hydrogen atoms, reacting with benzene.

We see how the replacement reaction discovered by the use of heavy hydrogen, discloses the nature of hydrogenation, which appears now to be a side reaction of the replacement reaction. Similar success may be expected in many other cases.

### THE LOWER REACTIVITY OF HEAVY HYDROGEN COMPOUNDS

The study of the hydrogen interchanges to which we originally turned, in order to derive some information on the energy differences between the ordinary and the heavy hydrogen compounds, has led us away from our starting point. We return now to the question raised at the beginning of this lecture, the lower reactivity of heavy water as compared with ordinary water.

It is yet uncertain to what extent the lower reactivity of heavy hydrogen compounds can be considered to be a general rule. But it is certainly a fairly widespread phenomenon. The possible interest of such lower reactivity is, of course, manifold. Hydrogen compounds which ordinarily are readily oxidised or otherwise decomposed, might become stable if the ordinary hydrogen is replaced by heavy hydrogen. Reactions might be

led into new paths or else their output might change considerably. Theory and practice would profit abundantly by such phenomena.

This lower reactivity of heavy hydrogen and of the compounds of heavy hydrogen can be explained by the theory of permanent energy with which I have dealt in the first part of my lecture. Indeed it was predicted from this theory, when there was still scarcely any experimental evidence for it.

The essential connection between permanent energy and reactivity is easily recognised. A molecule undergoes chemical reaction only if it happens to accumulate a certain critical amount of energy. The molecule has to wait until, in the course of the constant fluctuation of energy caused by heat motion, it happens to get an especially big share of energy equal to this critical energy. As soon as it has swallowed this, it goes to pieces : that is, chemical reaction.

Now suppose we have two molecules, one a compound of ordinary hydrogen, the other, the corresponding compound of heavy hydrogen. Let both molecules wait together, until by a fortunate fluctuation of thermal motion, they acquire the critical energy necessary for reaction. The ordinary hydrogen molecule is obviously in a better position in this competition, since it has a start on account of its greater permanent energy. The energy to be acquired by it is correspondingly smaller, and it will have an earlier chance to get this smaller quantity. It will, therefore, react before its competitor, the heavy hydrogen compound.

This is the reason for the lower reactivity of heavy water, and of other heavy hydrogen compounds.

## WATER WITH HEAVY OXYGEN

I have mentioned before that ordinary oxygen of the atomic weight 16 is accompanied by small quantities of a heavier isotope of weight 18. This heavy oxygen forms with hydrogen a heavy water of a kind quite different from "ordinary" heavy water. $H_2O^{18}$ in pure form would have about the same density as $D_2O$, that is 10 % above ordinary water density.

The two sorts of heaviness could be combined in "super-heavy" water, $D_2O^{18}$, which would have the density $1 \cdot 2$.

It is, however, much harder to prepare pure heavy oxygen than it is to prepare pure heavy hydrogen. Although the abundance of heavy oxygen in ordinary water is eight times higher than that of heavy hydrogen, it has not yet been isolated. The difficulty is that the two sorts of oxygens are chemically identical, and hence we have no convenient hold whereby to grasp the one, leaving the other back. The separation can be carried out only by physical methods which are comparatively ineffective.

The best physical method for the separation of isotopes is at present the " fractionated diffusion " of G. Hertz. By this method, Professor Hertz has succeeded in preparing about 300 mg. of water containing about 1 mg. of heavy oxygen. Professor Hertz gave us this sample and Dr. Szabo and the author have made the following use of it.

To the water we added a small quantity of metallic sodium, thus forming an alkaline solution. Then a few milligrams of amylacetate were treated with this solution until completely saponified. From the amyl alcohol produced by the saponification, the hydroxyl group was split off

in the form of water.  We examined this water, and found that *its density was normal.*

It follows that this oxygen does not come from the water used for saponification ;  it must, therefore, come from the oxygen of the ester-bridge.  Or, in chemical symbols :

$$Am\ O\ \ Ac + H_2O^{18} = AmOH + HO^{18}\ Ac$$

and *not*

$$Am\ \ OAc + H_2O^{18} = AmO^{18}H + HO\ Ac$$

This decides a question that had remained open since it was first raised by Van t'Hoff in 1899. Although the answer may only confirm what many chemists had surmised before, it still shows how useful heavy oxygen might become for the elucidation of the mechanism of oxygen reactions, like hydrolysis, oxidation, etc.

### HEAVY ISOTOPES OF OTHER IMPORTANT ELEMENTS

We have seen that the differences in the chemical properties of ordinary and heavy hydrogen are interesting, both in themselves and as a means of preparing pure heavy hydrogen.  But often the heaviness of the new hydrogen is used merely as a convenient " label " to mark the path which the hydrogen follows, when it becomes mixed and interchanged with other hydrogen atoms.  For this labelling purpose, the heaviness of heavy oxygen, $O^{18}$, has turned out to be just as useful a tool where reactions of oxygen are concerned. The same is obviously true for $N^{15}$ and $C^{13}$, with respect to the study of reactions involving nitrogen and carbon.*

* Labelling of atoms by isotopes was first introduced by Hevesy and Paneth in their method of " radioactive indicators." [See Hevesy and Paneth, Lehrbuch der Radio-aktivität. J. A. Barth, Leipzig (1923) p. 705].

Heavy hydrogen has a start over the other isotopes of the more important elements, because it was the first to be isolated. When we have the other isotopes at hand in sufficient quantities, they may well prove even more important than heavy hydrogen. All branches of chemistry will profit richly by such progress, but probably the greatest stimulus of all will be given to the chemistry of living matter when labelled carbon, hydrogen, oxygen and nitrogen atoms become generally available.

[M. P.]

# Friday, February 1, 1935

HARRY EGERTON WIMPERIS, C.B.E., M.A., M.I.E.E.
*Manager and Vice-President, in the Chair*

F. SIMON, D.Phil.
*Clarendon Laboratory, Oxford University : Formerly Professor
of Physical Chemistry at Breslau*

## THE APPROACH TO THE ABSOLUTE ZERO OF TEMPERATURE

SPEAKING in this building, with all its associations with low temperatures, a long introduction would be superfluous, but I would like to remind you of some data. In Fig. 1 a temperature scale is given in degrees Centigrade, and you see some important fixed points marked on it.

Now is this scale infinite at both ends or not ? We know that heat consists in the unordered motion of the smallest particles, the atoms or the molecules, and the intensity of this irregular motion rises with increasing temperature (in an ideal gas it is directly proportional to it). So it is evident that there will be no limit to high temperatures, as there is none to the intensity of the motion, but of course, there will be a lower limit to the temperature scale, at the point where the thermal motion stops altogether. This point is therefore with justification called the absolute zero. Though it has not been reached in experiment, and as we will see later on, it is by principle impossible to reach it absolutely, its position can be given with great accuracy, and it is found to lie at $-273 \cdot 1°$ C. A rational temperature scale has, therefore, to begin at this point. In this scale, the Kelvin

scale, the absolute zero of temperature is given by
the number zero, and any other temperature by
adding $273 \cdot 1°$ to the number of degrees Centi-
grade. On the right hand side of Fig. 1 the
temperatures are indicated in degrees Kelvin, and



| °C | °K |
|---|---|
| 100 | 373 BOILING P! $H_2O$ |

HELIUM REGION

| | °K |
|---|---|
| 15 | 288 ROOM TEMP |
| 0 | 273 MELTING P! $H_2O$ |

| | |
|---|---|
| 5 2 | CRITICAL P! |
| 4 2 | BOILING P! |

| -183 | 90 BOILING P! $O_2$ |
| -195 | 78 BOILING P! $N_2$ |

| -253 | 20 BOILING P! $H_2$ |
| -269 | 4 BOILING P! He |
| -273 | 0 |

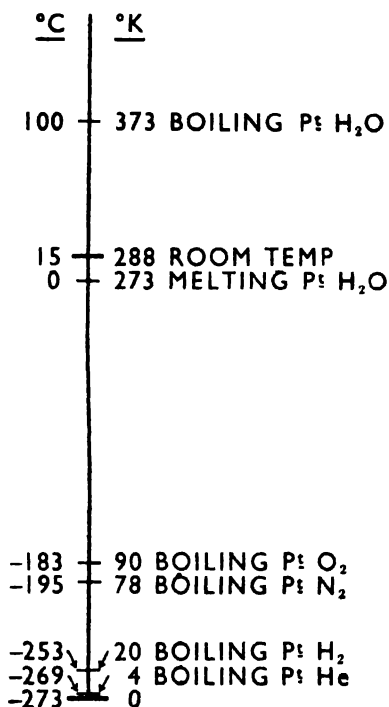| 1 0 | |
| 0 7 | KEESOM |
| 0 | |

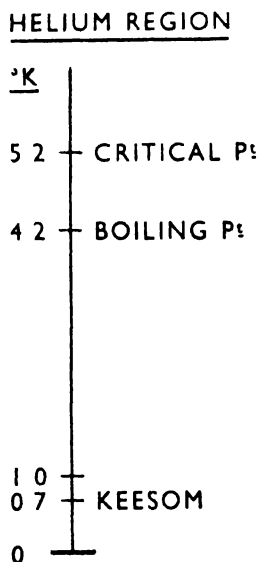FIG 1                    FIG 2

you see that the boiling point of the most volatile
gas, helium, lies about 4° from the absolute zero.
In Fig. 2 you see the liquid helium range given in
more detail. By reducing the pressure over the
liquid helium, one can easily get down to about 1°.
By improving the isolation, and using a huge

pump, Keesom succeeded in reaching $0\cdot7°$, which was the lowest temperature obtained until a few years ago.

I will begin by showing two experiments. The first is the production of a high temperature. It is a very simple experiment that you have all done yourselves. I need only switch on an electric lamp. By varying the current I get temperatures up to 2,500°.

In the second experiment I will generate a very low temperature, and that is much more complicated. While I could certainly perform the first experiment without any help, the second would be impossible without the kind help of Mr. Green,. Dr. Kürti, Dr. London, and the staff of the Clarendon Laboratory, who have all helped in its preparation. We will now liquefy helium, making use of a principle about which I will say something later. Now I should like to point out that we start at a temperature obtained with solid hydrogen (about 12°), that within the apparatus there is helium compressed to about 100 atmospheres at this temperature, and that the helium will be expanded into a balloon, where we will store it. During the expansion the helium will liquefy and reach its boiling temperature, $4\cdot2°$. As the helium is enclosed in a metal vessel, I will not be able to show it to you, but even if I could, it would not look a bit colder than liquid air. This simplified gas thermometer gives you a much better measure for the temperature. You saw that it pointed first to 12°, then, as the helium was expanded and this balloon filled, the temperature fell to about 4°. I can also show you another proof that we have reached the temperature region of liquid helium.

You have all heard of supraconductivity. This is the name given to the phenomenon discovered by Kamerlingh Onnes, that at certain temperatures some metals lose their electrical resistance altogether. The temperatures at which this happens are very low, they all lie below 10°. In a normal substance a current once induced would disappear very quickly, as its energy would be absorbed within about 1/1,000 of a second by its resistance. In the case of a supraconductor, however, there is no resistance, so that the current goes on flowing indefinitely.

In this lecture room, Professor McLennan* showed you this phenomenon of a persistent current in a lead ring, which was brought from Léiden immersed in liquid helium, so I need not speak about it in detail now. For the moment we will only use it to show that we really have the temperature of liquid helium. For this purpose a lead ring is fixed within the apparatus. Lead becomes supraconductive at about 7°, so that now at 4° it is already in the supraconducting state. We will now induce a current within this ring.†

Before the current was induced a magnetic needle was not affected. Now you see that when I hold it a little above the ring it points with its N pole towards it. This indicates that there is a S pole of the magnetic dipole corresponding to the persistent current. Bringing it below the ring, it naturally changes its direction.

I can show you the existence of this current in yet another way. I have here a small coil con-

---

* Proc. Roy. Inst. Vol XXVII, p 446   1932.

† This was performed by switching on a magnetic field, higher than the threshold value of lead at this temperature, so, with the field on, no persistent current flowed, reducing the field, a current is induced, which now persists as the substance is in the field zero.

nected with a galvanometer. When I bring it up to the apparatus, you see a ballistic deflection of the galvanometer. This is due to the cutting of the magnetic lines of force originating in the persistent current, and the magnitude of this ballistic deflection is a measure of the intensity of the current. So we have here a quantitative measure for the current, and we will verify at the end of the lecture that it is steady.

You see the great difference in outlay for the two experiments. For generating 2,500° you have only to connect an electric lamp to the mains; for generating −270° you have this decidedly complicated apparatus—and yet this is certainly the most simple one in existence. And we have not yet even taken into account that in the second experiment we did not start at room temperature, but at −260°. This temperature was obtained by hydrogen which we liquefied in the plant in the Clarendon Laboratory in Oxford, and the liquid nitrogen necessary for the hydrogen liquefaction was generated in yet another plant. For the first experiment, on the other hand, I could get the whole plant in my hand by taking a torchlight.

Now what makes this big difference ? Heat is due to the irregular motion of the smallest particles. So to heat a substance means to increase its energy, but still more to increase the internal disorder of its particles. To cool down a substance means to diminish its energy, but still more to increase the internal order of its particles. For example, a gas is in a very big state of disorder, the atoms flying in all directions about the space according to their kinetic energy, which is proportional to the temperature. Cooling it down,

it will first liquefy, and a liquid is already a much more ordered system than a gas. Cooling it still further, it will solidify ; now all the mean positions of the particles are given by a crystal lattice, and the system is in a state of nearly perfect order, only the thermal vibrations of the atoms around their mean positions being still a source of some disorder. Cooling down still further, even this disorder vanishes more and more.

So cooling a substance means bringing it into a state of order, and it is always much easier to make disorder than to establish order. Although you are certainly very familiar with this fact, I will illustrate it by an example. On this tray I have arranged these black and white balls in order. It is very easy now to establish a disorder by shaking the tray, but it is impossible to establish any sort of order again by shaking, or to speak more precisely, the probability of succeeding is extremely small. Of course, I can establish an order by selecting the balls with my hands, but in the case of a system consisting of atoms, this is obviously impracticable, and by principle impossible too. We have only macroscopic means at our disposal. This difficulty of creating order is just what makes the big difference. And this is why, even 10,000 years ago, men were able to generate very high temperatures—sufficient for melting metals—but the low temperature technique is of a quite recent date. Even the ice cream industry is fairly new, though the temperatures involved are not very much below room temperature.

To generate heat one must, of course, have energy at one's disposal, but to transform this energy into the disordered form of heat energy presents no difficulty at all. We saw this in the

electric lamp, but one can do it in a great many other ways ; for instance, in a candle it is a chemical energy we transform into heat energy, in the brakes of a car it is a mechanical one.

If our example with the balls were absolutely analogous, it would be impossible to generate a low temperature at all, but luckily it is not so, because, unlike our analogy where the order depends only on one variable, in a real physical system it depends on many more. The most important quantity it still depends upon is the volume. Taking the probability of finding an atom within a certain region as a measure for the state of order respective to the positions in space, it is quite evident that this probability decreases on enlarging the space the atom has at its disposal. So in a diluted gas we have a great disorder ; if it is compressed to a small space, its order is increased.

Now thermodynamics has given us a quantitative measure for this state of order of which we have to speak so much now, since it is necessary for understanding things later on. This measure is called the entropy. I will not trouble you with this quantity, as I know it is not a very popular one. I only want to remind you that it is a measure for the state of order, and that there is a law, namely the second law of thermodynamics, which tells us that within a closed system during any change the entropy can only increase, or at the best by making a reversible change (that means avoiding unnecessary disorder) it can remain the same. Speaking in our terms now, this second law means that in a closed system the state of order can only decrease or at the best remain constant.

Let us see now what this has to do with generating low temperatures. We will take a

cylinder with a piston, containing a gas, the whole system being perfectly isolated from its surroundings. The state of order in this system consists of two parts, one depending on the temperature, and the other on the volume. Compressing the gas, we increase the state of order corresponding to the volume. As the whole state of order must remain the same, the disorder due to the thermal motion has to increase, that means the temperature rises, and you all know that in compressing a gas it heats up, this heat being called the compression heat.

Bringing this system into thermal contact with its surroundings, it will cool down to the initial temperature, and so its disorder becomes smaller. But of course, that does not contradict the law I spoke of before, because heat is transmitted to the surroundings, so increasing the disorder of the particles there.

Now we will isolate the system again and pull the piston out. The part of the disorder due to the volume increases again. The whole state of order must remain constant, so the part of the disorder due to the temperature must fall, and that means the temperature itself falls.

That is a characteristic example, and one of the most important cases of how to generate a low temperature. Generally speaking, whenever one wants to lower the temperature, one must have a system in which the state of order can be changed by some external means. Then in the way described above, one is able to transmit a part of the original disorder in the system to the surroundings, and to cool it down, making the same change of this variable in the opposite direction after having isolated it from the surroundings.

Gases are the prototype of a disordered system, the existence of which is necessary for the procedure of generating low temperatures, and it is relatively easy with them to change this disorder by changing the volume. So practically all procedures for generating low temperatures were worked with gases until recently. Of course, one can make use of them only down to the temperatures at which they liquefy, or more precisely, as long as their vapour pressures have still practicable values. So in practice the generation of low temperatures and the liquefaction of gases have become practically identical conceptions, and every step towards a lower temperature has been marked by the feat of liquefying a gas with a lower boiling point than was previously possible.

I will not speak now of the development of the real procedures performed with gases in order to liquefy them, which, for technical reasons, have to be much more complicated than the example I gave you. I need only remind you of the names of Faraday, Cailletet, Pictet, Olzewski, Linde, Hampson, Dewar, Claude and Kamerlingh Onnes. You know that it is now relatively easy to get down to the temperature of liquid air, as the liquefaction of air has become important for industrial technique. But I want to mention that to cool by one calorie, at even such a relatively high temperature as that of liquid air, is already 500 times as expensive as to heat by a calorie above room temperature, for instance, by burning benzene. But at lower temperatures the difficulties increase enormously, so that the use of liquid hydrogen has been restricted to a few laboratories, and that of liquid helium to still fewer big specialised laboratories.

In recent years the study of the properties of matter at very low temperatures has become increasingly important, and so, of course, one has sought for ways of simplifying the low temperature technique. We have in the last few years developed a comparatively simple method for liquefying helium, which I have shown you already, and now I should like to say a few words about it. The procedure of expanding a gas in a cylinder, in the way already described, is very simple. But at very low temperatures it is difficult to realise technically a cylinder and a piston.

But one can overcome this difficulty in a very simple way which I will now explain. In Fig. 3 A we have a gas enclosed in a cylinder with a piston. Pulling out the piston, the gas will cool down. In Fig. 3 B you see the cylinder divided into two parts connected by a tube. Pulling out the piston now, the gas will cool everywhere, because this cooling is an homogeneous procedure. Let us cool now only the lower part to the low initial temperature at where we start (for example, to the temperature of liquid hydrogen if we want to liquefy helium); leave the upper part at room temperature, and pull out the piston again. Then the gas will cool down within the lower part, the same as if the upper part were at the low initial temperature too. The atoms in the lower part do not know whether the upper part is hot or cold and the atoms do not know either if there is a cylinder and piston outside. We get the same effect if I simply let it out by a valve, as in Fig. 3 C. I want to emphasise that the cooling arises within the cylinder, and not at the valve, as in the Linde process. The procedure described above has nothing to do with the Linde process, but it has more resemblance to the

Cailletet method. You remember, perhaps, that he let a gas expand in a glass capillary tube, and with some of the so-called permanent gases, he could then see a little dust of liquid drops, indicating that for some fraction of a second the temperature had fallen considerably.
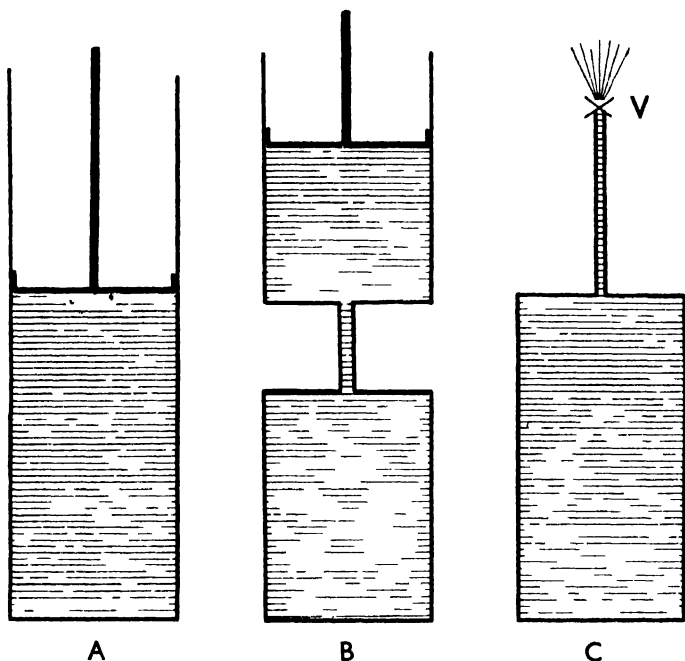


Fig 3 —Principle of the Expansion Method

Now starting at high temperatures, the cooling effects obtained in this way are very small. The chief reason for this is that a container for high pressures has, at room temperatures, a heat content that is always big compared with that of the gas. Starting at low temperatures, however, the situation changes absolutely. Firstly, one gets a much

bigger amount of gas into the container at a given external pressure, according to the gas laws, and secondly, the specific heats of all solid bodies drop with temperature, disappearing on approaching absolute zero.  So for instance, at a temperature of about 12°, 1 cu.cm. of helium gas compressed to 100 atmospheres has the same heat capacity as 1 kilogram of copper.  This means we can neglect the heat capacity of the walls altogether, and we have the advantage of working with mathematical walls.  Thus the efficiency of the procedure as I described it to you becomes very high, and working with suitable dimensions and a good isolation, it is easy to keep the low temperatures too.  For instance, under the conditions realised in this apparatus, about 60 per cent of the volume originally filled with the compressed helium remains filled with the liquid phase, and in the apparatus we generally use we can raise this efficiency still much higher.

In this way one could liquefy any quantity of helium.  But as the specific heats at low temperatures are so very small, only tiny amounts of liquid helium are really necessary for cooling down the apparatus and making measurements for a number of hours, if the apparatus is designed in a suitable way.  For example, in this apparatus we have liquefied about 50 cu.cm., which is sufficient to cool down the whole system and to work for about five hours.

In Fig. 4 you see a rough plan of the apparatus. Outside is the vessel D with liquid hydrogen ; inside, the space S that is evacuated.  Within this you see the container C that is first filled with compressed helium, and after the expansion with liquid helium.  Attached to this container is a gas
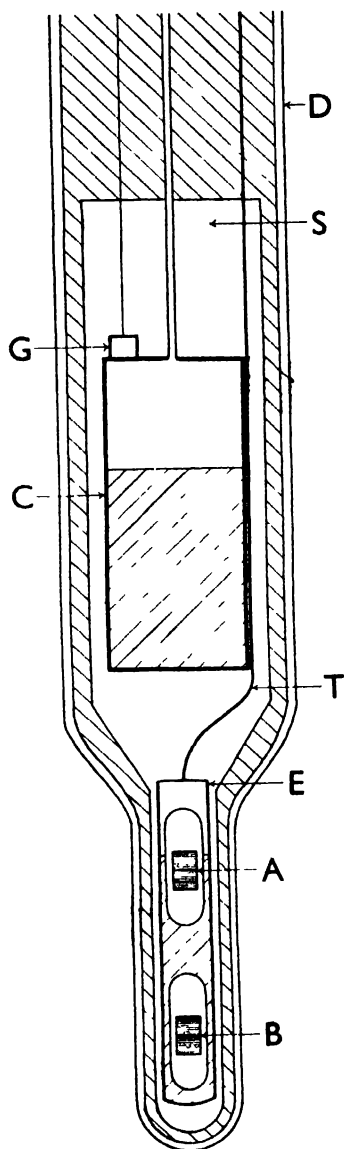
FIG. 4 —Simplified diagram of the apparatus.

thermometer G, the readings of which you saw before.

To cool the apparatus down further, one could reduce the pressure over the liquid helium, the temperature must then fall till it corresponds with the vapour pressure of the helium. For purely technical reasons we do not do this, but we have a second vessel E that we can fill with liquid helium by letting helium gas through the tube T. Then it condenses in the tube T where it is in contact with C, and drops down into the vessel. By pumping through T now, we reduce the vapour pressure, and therefore the temperature falls. This vessel and the surrounding Dewar vessel are of this peculiar shape because we will afterwards apply magnetic fields to some substance situated in the vessel, and of course, one can generate strong magnetic fields over a small distance more easily than over a big one. Now we will pump off the helium in this vessel, in order to reduce the temperature here, and the temperature will fall below 2° within a short time.

Before we consider the methods of generating still lower temperatures, I want to refer briefly to the measurements of these temperatures. You know that in general one measures temperature by the pressure of a diluted gas, using it as a substitute for an ideal gas. But at the lower limit of the liquid helium range the only existing gas, the helium, is stable only at a pressure so small that it is of no use for a thermometer. In this region one can use another phenomenon for measuring the temperature.

The thermometer that I want to speak of now is a magnetic one, and it depends on the fact that paramagnetic susceptibility is a function of tem-

perature. In a paramagnetic substance there exist little elementary magnets, which we will assume for the moment to be perfectly free to point in every direction in space. The thermal agitation has the effect of making the directions of these elementary magnets have a random distribution. Applying a magnetic field, it will try to turn them in its direction ; on the other hand, the thermal agitation tries to establish a disorder with respect to the direction of the dipoles. There will be a compromise of these two effects, and it is evident that the lower the temperature, that means the lower the thermal agitation, the more magnetised the substance will become. Calculating this numerically, one finds that in such a substance the magnetic susceptibility would be proportional to $1/T$. That is the famous Curie Law, which was derived primarily for paramagnetic gases, because there the single elementary magnets connected with the atoms or molecules are certainly perfectly free. At first sight one might think that within a solid body the condition of free elementary magnets could not be realised. But experiments, made chiefly in the Leiden Laboratory, have shown that there are some paramagnetic salts which follow Curie's Law down to the helium temperatures with great accuracy, and this means that the elementary magnets within them have a very high degree of mobility. The salts concerned are especially those of the rare earth and iron groups. I cannot go into a theoretical explanation of this behaviour, but I want to mention that it is in accordance with our theoretical ideas.

So it is easy to construct a thermometer with such a paramagnetic substance, using its susceptibility as a measure for the temperature. And

it is evident that a thermometer of this kind becomes very sensitive at low temperatures. For example, between 4° and 1° the susceptibility changes in the proportion 1 : 4. We have a thermometer of this kind in our apparatus. We measure the susceptibility by bringing a system of two coils around the place where, inside the apparatus, the salt is situated, and sending an alternating current through the primary coil ; the induced e.m.f. in the secondary depends on the susceptibility of the salt which is fixed within these coils. This induced e.m.f. is amplified in a little set (which the Cambridge Instrument Company have very kindly lent me for the demonstration) then it is rectified and sent to a galvanometer that shows the deflections on this scale. Of course, the system is compensated in such a way that the deflection zero corresponds to a susceptibility zero within the coils. Thus the deflections on the scale are a direct measure of the susceptibility, and as this changes with $1/T$, the deflections are proportional to $1/T$. At the moment the thermometer is showing about 2°.

Now we can continue considering how to approach nearer to the absolute zero. As I have mentioned already the lowest temperature reached by reducing the vapour pressure of liquid helium is 0·7°. At this point the vapour pressure is so small that it is in practice impossible to proceed further. A gas with a still lower boiling point does not exist. Very probably the new helium isotope, helium 3, discovered in the Cavendish Laboratory, will be more suitable for reaching low temperatures, if it can ever be obtained in sufficient quantities, but there will be no difference in the order of magnitude.

You may be wondering why we should bother to get still nearer to absolute zero, as it seems so difficult now to get down any further. What can still happen in this small region ? To answer this question, we have to ask another. When can one predict that something will happen in a certain temperature region ?

Let us assume that the phenomenon we are interested in is connected with an energy change of a certain quantity. Then the thermal agitation will have an influence on it when it itself reaches this order of magnitude. So we see that from this point of view, there is no sense in speaking of an absolutely high or an absolutely low temperature ; it is always necessary to compare it with the phenomenon in which we are interested. For instance, room temperature is a very low temperature if we look for the evaporation of diamond, because its heat of evaporation is very high (i.e., only at high temperatures is the thermal agitation big enough to push a carbon atom out of the crystal.) But room temperature is a very high temperature if we look for the evaporation of hydrogen, as its heat of evaporation is very small. So the question is, are there any phenomena connected with very small energy changes, that means, phenomena which will still happen at very low temperatures ?

If the atoms were only points possessing attractive or repulsive forces, then certainly nothing much of interest would happen within the new region. The thermal agitation would become smaller, but this would not give rise to any new phenomena. However, we know that although in the kinetic theory it was for a long time sufficient to treat the atoms as points with attractive and re-

pulsive forces, yet this is certainly not a complete picture. We know that the atoms are built up from nuclei and electrons. In general one is accustomed to find the effects of this complexity of the atoms only in gases at high temperatures, as most of them are connected with big changes of energy. It is true that at normal temperatures the effects originating in the complexity of the structure of the atom are not very striking in a solid body, but certainly some do exist. For instance, we spoke just now of the magnetic properties of some salts. If the atoms were only points with forces, they could not show any magnetic properties. These are due to the motion or the spin of the electrons, and here we have one effect of the structure of the atom. We have seen already that this effect becomes more and more striking as the temperature is lowered. We have also considered another phenomenon that would not have been possible if the atoms had only been points. A system of points could not show metallic conductivity. That is due to electrons split off the atoms within the metal, and we have seen that with these electrons, something happens only at very low temperatures, namely supraconductivity. Here some change takes place connected with an energy difference of such an order of magnitude that it becomes equal to the thermal agitation only at very low temperatures ; and I may mention that at present it is not known what exactly is happening in the metal. It would be very important to see whether, at still lower temperatures, all metals become supraconductive, that is, whether it is a general property of all metals.

So we see that it is of interest to extend our temperature range to lower temperatures, and we

will find later on that there are still more pheno-
mena that can be expected to take place below 1°.

But in order to go lower, how can we proceed ?
As we discussed before, to get a low temperature,
one must have a system at one's disposal which is
still in a big state of disorder, and one must be able
to change this disorder by changing an external
variable. At these temperatures we no longer have
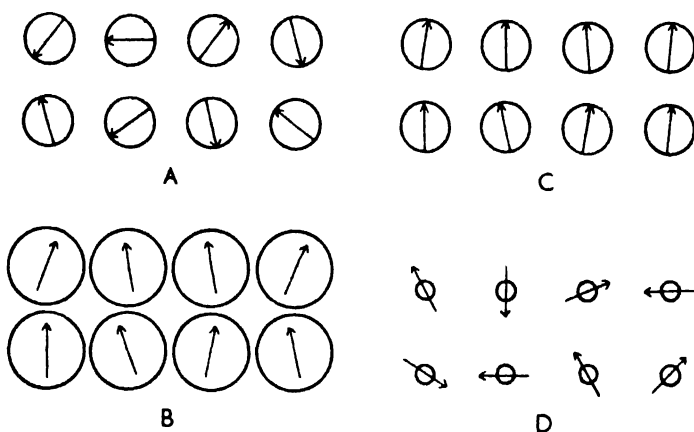gases. What other disordered systems still



Fig 5 —Principle of the magnetic cooling method

exist ? Well, we have just seen one, namely the
paramagnetic salt which still follows Curie's Law,
and about ten years ago Debye and Giauque
proposed using this for the generation of still much
lower temperatures.

To understand the principle, look at Fig 5
which represents a paramagnetic salt. The circles
represent the atoms arranged in a crystal lattice,
and the little arrows represent the magnetic
moments attached to every atom. Without a
field (Fig. 5 A) there exists, as we saw before, a

random distribution of the directions of the dipoles, so the disorder in this system consists of two parts; one is due to the distribution of the directions of the dipoles, and the other to the thermal motion of the atoms. (The diameters of these circles indicate the magnitudes of the vibrations and give also in this way a measure for the disorder to the thermal vibrations.) Applying the field now (Fig 5 B) it will try to order the directions of the dipoles. Making it adiabatic, that is, having the system isolated thermally from its surroundings, the whole state of order should remain constant, which means that the disorder due to the thermal movement must increase, in other words, the temperature rises, (corresponding to the compression of the gas in our former example.) Making heat contact with the surroundings the system cools down to the initial temperature, so that now the state of order has increased (Fig. 5 C). Now, isolating the substance from the surroundings, and taking the field away, the dipoles try to distribute their directions at random again, the disorder due to this increases, but as the whole state of order must be constant, the temperature must fall (Fig 5 D). (Corresponding exactly to the adiabatic expansion of the gas.)

Within the last few years this method has been used experimentally, and at nearly the same time Giauque in California, de Haas in Leiden, and Dr. Kürti and I began to work with it. We developed the technique so that it is now fairly easy to work with this procedure, and we can show you an experiment with it here in this room.

We will look once again at Fig. 4. There you see within the lower vessel two different para-
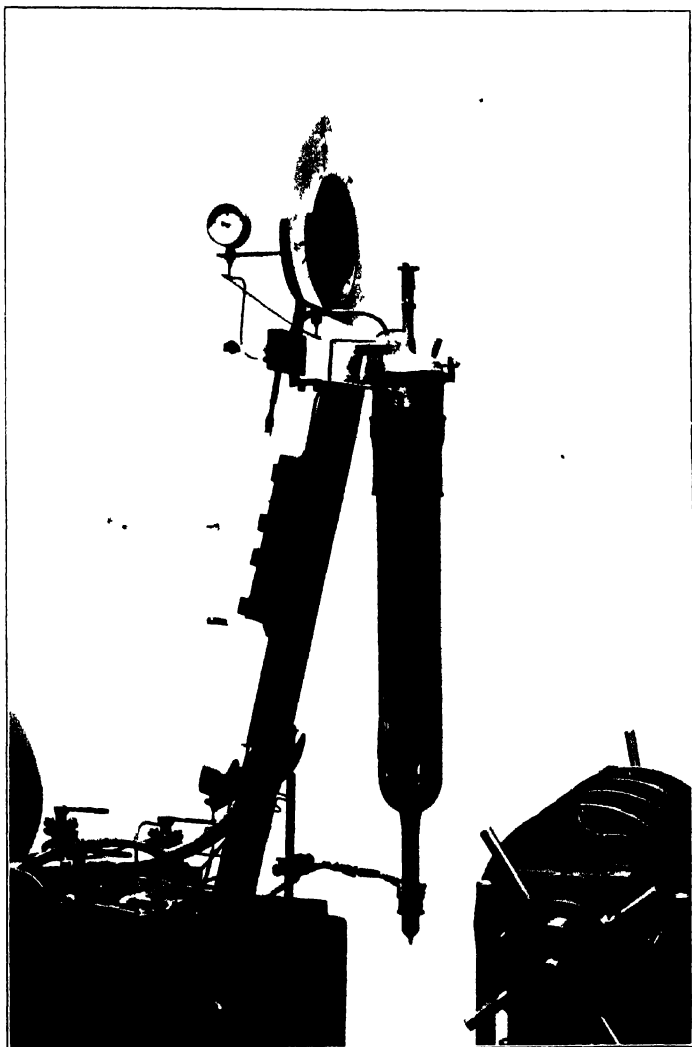
PLATE I — The Apparatus   In the middle the Dewar vessel containing the
liquid hydrogen in which the helium apparatus is immersed , at the top the
gas thermometer, indicating the temperature of the helium liquefier   On the
thin part of the Dewar vessel are the coils for measuring the susceptibility
of the salt , in the right bottom corner the magnet, which can be moved to
surround the lower part of the Dewar vessel   On the left, the balloon into
which the helium is expanded can just be seen

magnetic salts A and B, so that we can do two
different experiments. To carry out the cooling
by this method, one has to have the substance
in thermal contact with the surroundings when the
field is switched on, and isolated when the field is
switched off. We do this automatically by
suspending the substance in a little glass tube
closed at both ends and filled with about 1 cm.
helium at room temperature. This gas makes a
heat contact with the surrounding helium bath
during magnetisation. Switching the field off
the substance cools quickly, and the helium gas
has to condense on it, as the vapour pressure falls
rapidly with falling temperature. Choosing the
right dimensions and vacuum conditions, one can
in this way use the cooling substance itself as a
pump.

Now we will begin with an experiment* taking
first the upper substance, A, manganese am-
monium sulphate. We bring the magnet into
position and switch on a field of about 10,000
gauss. We have to wait now for about a minute
until the heat of magnetisation is carried away.
Next we remove the magnet and bring the coils for
the temperature measurement into position. The
substance has cooled down, the thermometer
points to about 0·25°. At the same time you
notice that the temperature keeps quite steady.

We will now make another experiment with the
lower substance, B iron ammonium alum, and the
same field. We have now reached 0·1°, and you
see that again there is hardly any change of
temperature.

* During the lecture a wire of the thermometer circuit broke, so
that the experiment could not be performed. However, it was
shown to a large number of the audience three quarters of an hour
later when the trouble had been repaired.

Now we will look at Table I, which gives the vapour pressures of helium at different temperatures. Although they have not been measured experimentally, these figures are very accurate, since we have all the necessary data at our disposal

TABLE I

VAPOUR PRESSURES OF HELIUM

| T | | p (mm.) |
|---|---|---|
| 1·0 | ... | $1·5 \times 10^{-1}$ |
| 0·7 | ... | $3·2 \times 10^{-3}$ |
| 0·5 | ... | $2·5 \times 10^{-5}$ |
| 0·3 | ... | $7 \times 10^{-10}$ |
| 0·2 | ... | $3 \times 10^{-15}$ |
| 0·1 | ... | $3 \times 10^{-31}$ |
| 0·05 | ... | $4 \times 10^{-62}$ |
| 0·03 | ... | $6 \times 10^{-103}$ |

for calculating the vapour pressures according to the second law of thermodynamics. You see that at $0·1°$ the vapour pressure is $10^{-31}$, and at $0·25°$ it would be about $10^{-12}$, so that there is practically no gas which could transmit heat from the surroundings. At $0·5°$, however, we have a vapour pressure $10^{-5}$, which is no longer negligible. In this case the substance would warm up much more quickly, and so we have the paradox that it is much easier to keep a temperature below, let us say $0·3°$, than one above this temperature.*

At these very low temperatures one can get any isolation one likes, for instance, at $0·03°$ we have a vapour pressure of $10^{-102}$. The surroundings of the substance are at a temperature of about $1°$,

* Of course, one could establish a vacuum by means of pumps, but at low temperatures the helium is absorbed in big amounts on the walls, and it would take a very long time to obtain a sufficiently high vacuum.

where the radiation is certainly negligible, and one can make the suspension so that very little heat is conducted to the substance. Thus we have really no difficulty in keeping temperatures as long as we like, even working with very small amounts of substances.

You saw that using different paramagnetic salts, we reached different final temperatures, which means that all substances are not equally good for this method. Of the many substances we investigated, iron ammonium alum was found the most suitable, and with it we have got down to about 0·04°, using a field of 14,000 gauss. De Haas reached 0·015° with this procedure, using potassium chromium alum, and having the huge magnet of the Leiden Laboratory at his disposal.

I may remind you that the temperature of a material body in the interstellar space cannot fall below 2° or 3° K., as it always has to be in equilibrium with the stellar radiation. So you see that in this case we can realise in the laboratory a lower temperature than we can find in nature, and we can surpass the conditions found in nature in still another way. We will look once again at the table of the vapour pressures of helium, the most volatile gas existing. In the interstellar space there is a vacuum of about $10^{-22}$ cm. Hg. You see that we have already reached this pressure in a space surrounded by a body at a temperature of about 0·15°, even when it is filled with the most volatile gas. At 0·03° the pressure would be so small that in the whole Galaxy we would not find one single atom in equilibrium with it. So, in the directions of low density and low temperature, we can surpass in the laboratory the conditions found in nature, whereas, in the opposite direction it is extremely

unlikely that we will ever reach the high temperatures and big densities to be found in the stars.

There is not much point in generating low temperatures simply for the fun of playing among the low figures. How can one make investigations on substances other than the paramagnetic salts ? The most simple way is to press the paramagnetic salt and the substance to be investigated together, so that they form a solid pill. In cooling down the salt, the substance cools with it. In this way we examined a lot of metals to see whether they became supraconducting or not, and three new supraconductors were found in the new region. Some other metals, however, did not become supraconducting down to $0.05°$.

At the same time these measurements showed us still another thing. Cooling an additional substance with the paramagnetic salt and seeing how far the temperature is lowered in demagnetising, compared with the temperature reached in demagnetising the pure salt, one can measure the specific heat of the additional substance, and see whether anything happens within the new temperature region. If there is any change of energy within the substance that is equal to the thermal agitation in the new region, then it should be seen in these specific heats. And there is a very definite thing that must be expected to happen in this region. As is known from the analysis of the atomic spectra, there is an interaction of the magnetic momenta of the nucleus with that of the electrons, chiefly the valency electron. As a result of this, the ground state of the atom is split up, and the energy difference between the different levels is of such an order of magnitude, that the new temperature region is characteristic

of it.   We have already got some results in this direction, but it would be proceeding too far now to go into details.

You saw that within this new temperature region there really are things that still happen, and it is not only a game with low temperature figures. One can also say that it will be both necessary and interesting to investigate temperatures lower than those yet reached.

What imposed the limit to the temperatures reached above ?   It should be possible to get down to any temperature if the dipoles in the salt were perfectly free, because then, even as near to absolute zero as one likes, there would be a random distribution of the magnetic dipoles, and we would always still have a means of lowering the temperature.   But here the same thing happens as in the case of gases.   If we had still a gas at these very low temperatures, we could certainly use it for lowering the temperature, but no more gas exists, because there are forces between the atoms that cause the system to pass over to a state of order—the crystal—and there is no longer any disorder that we could make use of for lowering the temperature.   In the paramagnetic salt the same thing happens, only at temperatures about 100 times lower.   In them there also exists interaction forces which have the effect of establishing an order within the system without an external magnetic field, and so at these temperatures the paramagnetic salts cease to be of use.   These temperatures differ, of course, from substance to substance, and we saw in our experiments that the manganese ammonium sulphate passed over into this state earlier than the iron ammonium alum.   Of course, one could get a bit further

by using still stronger magnets, but I think the practicable limit has been reached with the dimensions of the Leiden magnet.

A more hopeful way seems to be to work in two stages. That means, first work down to about 0·05°, and then starting at this temperature go still lower with a new procedure, like the cascade for liquefying gases, which was used so much in the past. Of course, to get to appreciably lower temperatures for the lower stage it will be necessary to find a substance in which these inter-action forces, which tend to bring the system into an ordered state, are still smaller than within the substances hitherto used. I have already said that the interaction forces between the nuclear moments and their surroundings are very small, and in the second stage one will have to try to work with a substance that exhibits nuclear paramagnetism.

But even here there are some interaction forces, and so this will only work for a bit of the way to absolute zero. And with all other phenomena that may still happen in the new temperature re-gion, it will certainly be the same. There exists a law, Nernst's Theorem, also called the third law of thermodynamics, which is confirmed by all ex-periments. It postulates that at absolute zero all substances are in a state of perfect order, or in other words, that the state of lowest energy must be a state of perfect order. And we know now that this means that it will be impossible ever to reach the zero of temperature absolutely. But this does not mean that one cannot get below a certain limit, say 1/10,000°. It will be possible to reach any finite temperature, be it as small as you like. But the technique of reaching such a temperature

will always be dependent on finding a pheno-
menon, connected with only a very small energy
change, happening within a system. And so
you see that this last degree, or as we may say
now, the last 1/100° to absolute zero, though
absolutely very small, stretches in reality an
infinite distance before us. And this infinity is
not an empty one, but one that is filled with
phenomena worth investigating.

[F. S.]

## Friday, February 22, 1935

CLIFFORD COPLAND PATERSON,
O.B.E., M.I.E.E., F.Inst.P.
*Manager and Vice-President, in the Chair*

A. O. RANKINE, O.B.E , D.Sc., F.R.S., M.R.I.
*Professor of Physics in the Imperial College of Science and Technology*

## SOME EXPERIMENTS IN GRAVITATION AND MAGNETISM

ALTHOUGH I recognise fully that Einstein's view of gravitation has superseded Newton's, in that it is unnecessary to attribute to an attractive force the accelerated motion of an apple falling towards the earth, I shall in this lecture, merely for convenience of expression, still speak of such forces. Falling bodies constitute our most intimate manifestation of gravitation, and the most usual means of studying accurately this terrestrial gravitation is to observe the motion of a pendulum, the bob of which rises and falls in a modified manner. Thus we find that the same pendulum beats at slightly different rates in different localities, in particular more rapidly near the poles than at the equator, implying that bodies are pulled more strongly by the earth the greater the latitude. In other words the weight of a body varies from place to place. We do not so often recognise the fact that the weight of a body at a fixed point on the earth's surface changes also with time, although this is implied by the explanation of the tides. The moon exerts small gravitational forces at the earth's surface, and these change in direction as

the moon moves in its orbit.    It is only recently
that these effects have been measured directly, by
means of that remarkably accurate pendulum, the
Shortt clock.    This lunar variation of gravity is
very small, and we could not expect to feel our
own changes of weight ;  but the new liner, the
*Queen Mary*, with the moon overhead, does in fact
weigh 20 lbs. less than when the moon is on the
horizon.

    It is not, however, with terrestrial gravitation,
nor planetary and stellar gravitation, that I wish
to deal this evening.    My chief purpose is to
demonstrate the universality of Newton's law by
exhibiting the attraction between much less mas-
sive bodies.    The experiment is not easy because
the forces operating are so very small.    Let me
illustrate this by asking you to imagine these two
spheres to be made of solid gold, chosen on
account of its great density, and its consequently
relatively large gravitational effect.    Each sphere
is five inches in diameter, and the distance between
their nearest points is one inch.    The gravita-
tional force pulling them together is only the
weight of one-eighth of a milligram, or that of one
of the square scraps of cigarette wrapping foil
(one-tenth inch square) which you see projected on
the epidiascope, with some postage stamps for
comparison of size.    We cannot hope to be able to
detect so small a force, unless somehow we can
eliminate the effect of the enormously larger force
with which the earth pulls each sphere.    If only
we could have a gravitational shield with which to
surround our spheres it would be much easier, for
then they would be free to move under the
influence of their mutual gravitation alone.    As
with the occupants of Mr. Wells's Cavorite enclo-

sure in their supposed journey to the moon, they would approach one another until they touched. The vigour of the bump and the magnitude of the attracting force may be judged from the fact that the approach through the separating inch would occupy a quarter of an hour.

The credit for solving the problem of eliminating the effect of terrestrial gravitation belongs to the Rev. John Michell, who, towards the end of the eighteenth century, devised the torsion balance, which, although not a gravitational shield, enabled a precise discrimination to be made between the strong forces due to the earth and the feeble ones it was desired to measure. The principle is very simple. Two equal masses are attached to a beam which is suspended horizontally by a thin wire attached to the centre, as in this model before you. In such a system small horizontal forces applied to the masses are able to twist the suspension without interference by the earth's gravitation, which acts vertically and plays no part in determining the rotation of the beam. With fine suspensions exceedingly small horizontal forces can be made evident, and the system has since found wide application in a great variety of delicate measurements. Michell died before he could apply it to the end in view, but soon after Cavendish succeeded in measuring with considerable accuracy the gravitational pull between 2-inch diameter lead balls at the ends of the beam, and 12-inch diameter lead balls placed near them. Other measurements followed, culminating in those of Boys nearly fifty years ago, when, with the refinements his genius introduced, he determined the gravitational constant with very high precision.

Actually Boys demonstrated such gravitational

attraction in this theatre in 1889, so that the experiment I am about to perform is not the first of its kind that has been shown here. It is, however, an experiment worth repeating, especially as I am using for the purpose a special form of portable torsion balance due to Eötvös, whose work on gravitation, although contemporaneous with that of Boys, is not so well known in this country. In this balance the two suspended masses are at different levels. Each consists of about $\frac{1}{3}$ ounce of platinum, the one attached to one end of the beam, and the other suspended about 1 foot below the other end. The whole is housed in a triple metal enclosure to protect the suspension from air currents. The motion of the beam, normally observed by a telescope incorporated in the apparatus, is for our purposes indicated by the spot of light on this scale, the light coming from a mirror attached to the beam. The 40 lb. lead sphere which you see here has been for some hours close on one side of the lower mass. You will notice that the light spot is in the same position as it was at the beginning of the lecture. Now, as my assistant has transferred the lead sphere to the other side, the light spot begins very slowly to move—slowly because the attractive force between the sphere and the suspended mass is so small, but surely enough. If we waited long enough—about half an hour—the spot would become stationary again in a new position, enabling us to deduce the full rotation of the beam and the gravitational force to which this is due.

The Eötvös torsion balance, being a portable and well-protected instrument, can be used out-of-doors. It has been applied very extensively to

measure the small changes of terrestrial gravity due to buried structures of greater or less density, with which mineral deposits of economic importance, such as oil, are associated, thus facilitating detection. For this purpose it is at least a thousand times more sensitive than the most accurate pendulum, and it has the additional advantage that, unlike the pendulum, it does not respond to the lunar changes of gravity to which I have referred. It discriminates between spatial and temporal changes of gravity, and its indications can thus be assigned wholly to terrestrial anomalies. I have been trying in recent years to secure the same advantages with regard to magnetic field measurements, so far not successfully. But this magnetic analogue of the gravity torsion balance, although not yet adapted for field use, has proved a very sensitive and useful instrument for certain magnetic experiments in the laboratory. I have such an instrument here, but before demonstrating its working I must say something concerning the properties of materials with respect to magnetism.

The forces operating in magnetism are not of the same general nature as in gravitation. Magnetism is not an inherent property like gravitation ; it can only be imparted to bodies under suitable excitation. For example, we can make a rod of iron into a magnet by placing it in a solenoid through which an electric current passes ; but most of the magnetism vanishes when the current stops. If we do the same with a hard steel rod, the magnetism produced is more permanent, and we can use this manufactured magnet to excite magnetism in other bodies brought near to it. A few materials—iron, steel, cobalt, nickel and certain alloys, are very easily magnetised, and may

be made into very strong magnets. For measuring the magnetic forces between these no very sensitive apparatus is needed. Observe, for example, these two cobalt steel magnets on this inclined plane. I have their similar poles directly opposite one another, and you see that the strong force of repulsion prevents the rolling magnet from reaching the one at the bottom—it rebounds instead and rolls partly up the plane again. Or in this little glass cell of which there is a projection on the screen. The small magnet at the bottom repels so strongly the one I now drop in that this floats above the other—the repulsion is equal to its weight (about 3 grams) even when the magnets are a quarter of an inch or so apart.

But these large forces are confined to the so-called ferro-magnetic substances of which there are few, and iron the chief. To demonstrate the magnetism which can be imparted to the vast majority of materials, we need again to be able to measure very small forces, and this enables us to divide substances into two groups—paramagnetic and diamagnetic. To demonstrate this I first employ a method virtually the same as that used here by Faraday. The simplest way of expressing the distinction is to describe as paramagnetic a body which is attracted by the pole of a magnet, and as diamagnetic one which is repelled. Here I have one of the pole-pieces of an electromagnet, and near it I can hang by a long suspension small specimens which I wish to test. The pole and the specimen are projected on the screen so as to be more clearly seen. This piece of platinum is attracted when the current is passed to create a strong pole—it is paramagnetic. Graphite, on the contrary, is repelled, being diamagnetic. If we

use brass there is a strong attraction, but this is a spurious effect due to free iron in the brass. It is extraordinarily difficult to procure brass which is free from iron, and this precludes the use of this otherwise convenient metal in the construction of really delicate magnetic instruments. Fortunately we find a useful substitute in aluminium, which, as you see, is only very slightly paramagnetic, in spite of the rather large amount of iron it contains as an impurity. Its virtue lies in the fact that it alloys with the iron forming a compound which is not ferro-magnetic, whereas in copper the iron remains free and thus displays its ferro-magnetism to such an extent as to mask the diamagnetism which pure copper would exhibit. I have here a second aluminium cylinder into a small hole in which I have introduced one per cent of free iron—about as much as the metal already contains in combination, and the strong attraction due to this iron is now evident.

So far we have been using a very strong magnet pole for our tests. Can we detect the forces due to a comparatively weak pole? Can we, for example, deflect visibly this compass needle by holding near it a piece of graphite? We cannot. Here we are confronted with a problem like that in gravitational measurements —the control exercised by the earth, this time by its magnetic field, is much too strong. We can escape this hindrance by having recourse again to the torsion balance principle, using a small magnet, instead of simply a mass, suspended from one arm of the beam. I have such a system here. I have described it fully elsewhere*, and all I need point out now is that the apparatus constitutes a considerable

* Proc. Phys. Soc. Vol. XLVI, p. 391. 1934.

improvement in the methods of delicate magnetic measurement. The example before you is the first model, made of brass before the objections to this metal and the comparative virtues of aluminium were recognised. Consequently it has by no means the full sensitivity attainable, yet you see that the rotation of the beam, as indicated by the reflected light-spot, is very large when a piece of graphite or bismuth is brought near the suspended magnet, which is a small one weighing less than ⅓ ounce ; and even a test tube of distilled water indicates the small repulsion due to its diamagnetism.

You will observe that aluminium gives a curious effect, which, indeed, was also noticeable in the Faraday experiment. At first it repels the magnet, but eventually attracts it. This temporary repulsion is due to the development of eddy currents in the metal on bringing it near the magnet, but when the relative motion ceases, and the currents dissipate, the attraction due to the paramagnetism of aluminium supervenes. If this experiment could be carried out at the very low temperatures with which Dr. Simon dealt here some weeks ago, the repulsion would be the only effect we should observe, because the eddy currents would continue indefinitely in the supraconducting metal. Indeed it is known that supraconducting metals do display the maximum diamagnetism theoretically possible.

Although the apparatus is very sensitive to outside magnetic influences which are near to it, so that I have to remove my keys and other iron articles on my person in order that my approach may not disturb it, it fulfils expectations in being irresponsive, like the gravity balance, to variations

with time of fields originating from distant sources, *e.g.* the diurnal variations of the earth's field and even magnetic storms. Only spatial distortion of magnetic field is measured—that brought about by changes of magnetic distribution in the immediate neighbourhood. A uniform magnetic field does not affect it ; the system is, in fact, completely astatic. Accordingly for effective work at the greater sensitivity I have indicated one needs to control only a relatively small surrounding area as regards the exclusion of iron and other sources of disturbance of the earth's natural field. I have not such control this evening, as regards either the iron and steel in the structure of the theatre or—not having dared to have you all searched before entry—that you carry on your persons. I am therefore not able to show, as I had at first hoped, a newly constructed apparatus, built mainly of aluminium, which exhibits quite definitely the slight paramagnetism of air in the gaseous state. I can only show instead the rather strong paramagnetism of liquid oxygen, which you observe, as Dewar first demonstrated here, is drawn up to the poles of the electromagnet.

I am, however, convinced that we now have the means at our disposal of making accurate measurements of the magnetic properties of gases even in weak magnetic fields. There is no apparent reason why there should not be attained in such measurements the same high precision as Boys achieved in gravitation. Only we must have conditions of isolation and quiet like those he found in his cellar in Oxford. To such isolation I hope soon to retire.

[A. O. R.]

## Friday, March 29, 1935

LORD EUSTACE PERCY, P.C., M.A., M.P.
*President, in the Chair*

THE LORD RUTHERFORD OF NELSON,
O.M., F.R.S, M.R.I.
*Professor of Natural Philosophy*

# THE NEUTRON AND RADIOACTIVE TRANSFORMATIONS

[ABSTRACT]

THE discovery by Dr. Chadwick of the existence of a new type of particle devoid of electric charge, called the neutron, has proved of great importance to our knowledge of the structure of atomic nuclei. Although only three years have passed since the first proof that fast neutrons are liberated when the element beryllium is bombarded by *a* particles, our knowledge of the neutron and of its remarkable potency in causing the transmutation of many elements has advanced with great rapidity. Researches on the properties of the neutron have been carried out in many countries and a great mass of important data has now been accumulated. This discovery has not only proved of much interest to the experimenter but also to the theoretician, and it is now generally believed that the neutron and the proton are the primary constituents in building up the structure of the nuclei of all elements.

The mass of the neutron is known to be about the same as that of the proton, but it is of great importance to fix their relative masses with accuracy, for some are inclined to believe that a

proton may be converted into a neutron and vice-versa under the conditions existing within a nucleus. Chadwick and Goldhaber have devised a method which seems likely to fix the mass of the neutron with certainty. They have found that the nucleus of heavy hydrogen of mass 2, can be broken up into a proton and neutron by the action of $\gamma$-rays of high frequency. If the quantum energy of the $\gamma$-ray is known and the energy of motion of the resulting particles, the mass of the neutron can be deduced.

Neutrons can now be produced in large numbers in a variety of ways. Bombardment of elements like lithium and beryllium by the ions of heavy hydrogen gives streams of neutrons, but the largest effects are observed when heavy hydrogen is bombarded by its own ions. The speed of emission and number of the neutrons depends on the nature of the transformation. Feather first showed that fast neutrons were very efficient agents in causing the transformation of nitrogen and oxygen nuclei. Fermi, Segrè and others have found that neutrons are able to produce transformation effects in a large majority of the elements.

When neutrons, produced by the action of $\alpha$-rays on beryllium, are used to bombard matter, in many cases a radio-active substance is produced which emits fast electrons and loses its activity with time according to the same laws as a natural radio-active body. More than 50 of these radioactive bodies have now been found, each losing its activity at a characteristic rate. In general it appears that the neutron, in consequence of its absence of charge, is able occasionally to enter the nucleus of an atom, causing a violent transformation, often accompanied by the expulsion of a proton or

α–particle. A new unstable element may be formed which breaks up gradually with the emission in most cases of negative electrons. In this way we have been able to show that many new unstable isotopes of the elements can be formed and studied, in addition to the stable ones already known. The method of detection of these unstable elements is exceedingly sensitive. A Geiger-Müller counter is generally used and this responds to each individual electron which passes through the counting chamber. The number of particles entering the chamber each minute can be automatically counted. Fermi and his collaborators have shown that the amount of transformation can in some cases be increased about 100 times by slowing down the neutrons. This can be effected by passing the neutrons through substances like paraffin or water, which contain hydrogen. In a head-on collision with a hydrogen atom, the fast neutron is nearly brought to rest. Under such conditions it more readily enters the nuclei of some elements and a much stronger radioactive element is produced.

In general, the neutron is the most efficient agent so far discovered for producing the transformation of matter. In many cases each neutron entering the material is captured, and leads to a disintegration of an atom of the element under examination.

Experiments were shown to illustrate the properties of neutrons and their power of producing new radioactive bodies. The ease with which the neutrons pass through lead and the rapidity of their absorption in elements like cadmium were demonstrated, and the production of radioactive silver

and fluorine was shown, also the increase of the effect when the neutrons were slowed down by their passage through paraffin. The disintegration of the nucleus of heavy hydrogen into a proton and neutron by the action of penetrating $\gamma$-rays was also shown.

[R.]

# Friday, November 15, 1935

SIR ROBERT ROBERTSON, K.B.E., M.A., F.R.S.
*Treasurer and Vice-President, in the Chair*

ROBERT ROBINSON, M.A., D.Sc., LL.D., F.R.S.
*Waynflete Professor of Chemistry, Oxford University*

## THE SYNTHESIS OF A NATURAL COLOURING MATTER

THE object of this discourse is not so much to give a complete account of a group of natural products, as to describe, by following one example through all its stages, the methods used by organic chemists to elucidate the molecular structure of complex carbon compounds. For this purpose the bright red pigment of the scarlet pelargonium has been selected. This colouring matter is not unique to the scarlet pelargonium, but is also found in other flowers and blossoms, for example in pink cornflowers and in certain varieties of dahlia. The necessary stages of investigation are three, namely, isolation and characterisation of the pigment as a homogeneous chemical individual, analytical and degradative experiments affording clues to the molecular structure, and finally synthesis of that molecular edifice by unambiguous methods. These latter two modes of study are entirely independent, and if the conclusions reached by both methods are identical, the validity of the analytical arguments is established.

First, then, we must consider the isolation of the pure colouring matter. The possibility of this was first indicated in 1903 by Griffiths, who

extracted the petals with 90 per cent alcohol, and after several further manipulations obtained coloured crystals which he showed to be free from nitrogen and sulphur. The experiments were not carried any further. In 1905 the botanist Molisch demonstrated the presence of crystalline pigments *in situ* in certain flowers, and he also devised an interesting method for the extraction and crystallisation of the colouring matters on the small scale. The scarlet pelargonium gives good results, and these can be demonstrated after covering the petals with a few drops of acetic acid and placing between two glass slides. On slow evaporation the pigment extracted by the acetic acid separates in microscopic needles at the boundaries of the liquid [demonstration of the crystals by micro-projection, × 100]. This experiment of Molisch was made the basis of a large-scale preparation by Grafe in 1911 ; he used no less than 28 kilograms of fresh petals, and eventually isolated 10 g. of a beautifully crystalline pigment, as well as 15 g. of amorphous material. Unfortunately Grafe misinterpreted his results, but he was the first to obtain the crystalline pigment of the pelargonium in a practically homogeneous condition.

Following a pioneering paper by Everest and Willstätter in 1913 on the colouring matter of blue cornflowers, Willstätter and Bolton, in 1915, were able to describe the isolation of the colouring matter of *pelargonium zonale* (var. Meteor) in a comparatively simple way. The dried petals were found to contain 7 per cent of the pigment and 60–65 per cent of this amount could be isolated in a pure form. The extract of the petals with 90 per cent alcohol becomes colourless on keeping,
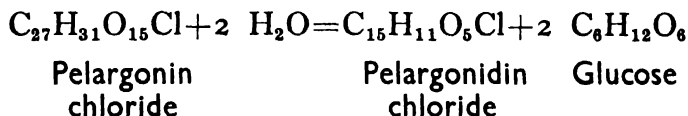
but the colouring matter is not thereby destroyed, and can be regenerated by the addition of a strong acid to the solution. [Demonstration : regeneration of the colour by addition of hydrochloric acid or tartaric acid]. The underlying chemistry of this transformation need not be discussed in detail, but it is apparent that the red colouring matter may be a compound of a base with an acid, and the isolation process employed by Willstätter was founded on this hypothesis. The acid which functions in the flower itself is probably tartaric acid, a relatively considerable amount of which can be isolated from the pelargonium. Willstätter, however, worked with the chloride in this as in other cases. To 1000 c.c. of the alcoholic extract he added 20 c.c. of 20 per cent alcoholic hydrogen chloride and the pigment was then precipitated by means of 2000 c.c. of ether. The sticky precipitate could be directly crystallised from aqueous methyl alcoholic hydrochloric acid, and the pure substance was eventually obtained in the form of glistening, very slender, orange-red needles.

Elementary analysis showed that the empirical formula was $C_{21}H_{31}O_{15}Cl, 4\,H_2O$. The substance was found to lose the four molecules of water of crystallisation on exposure to a vacuum at the room temperature. This *pelargonin chloride* is the subject of our study ; it satisfies all the requirements of homogeneity, and a specimen of it is therefore an assemblage of identical molecules characterised by a structure which must be determined.

Progress in this direction is made by decomposing the pelargonin chloride with the formation of less complex substances, some of which can be

recognised and are of known constitution. These fragments represent blocks of the complete structure and the only problem is how to fit them together.

In the case of pelargonin, the molecule is easily resolved into three parts, two of which are identical and recognisable as the sugar glucose. On boiling with 10–15 per cent hydrochloric acid, pelargonin chloride suffers a decomposition represented by the equation :—

$$C_{27}H_{31}O_{15}Cl + 2\ H_2O = C_{15}H_{11}O_5Cl + 2\ C_6H_{12}O_6$$

| Pelargonin | Pelargonidin | Glucose |
| chloride | chloride | |

There is no great colour change during this process, because the part of the molecule responsible for the colour suffers no destruction and remains intact in the form of pelargonidin chloride ; the sugar is colourless. The formation of two molecules of glucose from each molecule of pelargonin chloride is deduced from the quantitative relation of pelargonin used to glucose obtained, and is confirmed by the composition of pelargonidin chloride and the fact that glucose is the only sugar formed.

[Pelargonin chloride with sugar in combination is much more readily soluble in very dilute hydrochloric acid than in *iso*amyl alcohol. With a mixture of the solvents the distribution ratio can be observed after shaking and it is seen that almost all of the pigment remains in the aqueous layer. A somewhat greater proportion passes into butyl alcohol, and this distribution ratio can be measured under standard conditions and is characteristic of the substance. Pelargonidin

chloride passes completely into *iso*amyl alcohol from an aqueous acid solution.]

Evidently pelargonin is a diglucoside of pelargonidin, and the hydrolysis of the glucosidic pigment is analogous to that of sucrose by means of dilute acids. [Demonstration of appearance of cupric reducing power after hydrolysis of sucrose by dilute hydrochloric acid and of potato-starch by the amylase of saliva.]

The glucosides are widely distributed in Nature, and their chemistry is well understood; they are compounds in which glucose, $C_6H_{12}O_6$, or $C_6H_{11}O_5.OH$, appears in the form $C_6H_{11}O_5.O.R.$ where R is a residue of a substance R.OH. The HO-group of glucose, implicated in this combination, is that which is responsible, when free, for the cupric reducing power of the sugar. There are four further hydroxyl groups and an oxygen atom in a ring, and the complete structure for a $\beta$-glucoside, the most common natural type, is
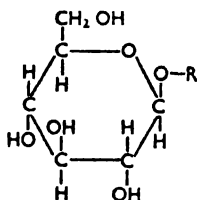


The diagram represents the arrangement of the atoms in space, as it has been ascertained that the OH groups lie alternately above and below the plane of the $C_5O$ ring of atoms.

Thus existing knowledge is a guide to the probable mode of attachment of the glucose residues to the pelargonidin nucleus but two alternatives are nevertheless apparent. The glucose groups may be line-ahead as in the scheme : $(C_{15}H_{10}O_4Cl)$

$O.C_6H_{10}O_4.O.C_6H_{11}O_5$, or they may be attached to different parts of the pelargonidin molecule as in the scheme : $C_6H_{11}O_5.O(C_{15}H_9O_3Cl)O.$ $C_6H_{11}O_5$. The first view represents a common arrangement but the second possibility was eventually proved to be applicable to the present case.

The detachment of $C_{12}$ from $C_{27}$ greatly simplifies our problem and the next stage is obviously the dissection of pelargonidin chloride.

Here more brutal methods must be used and on fusion with caustic potash the molecule is completely broken up. Two derivatives of benzene can, however, be salved from the wreck and these, namely phloroglucinol and parahydroxybenzoic acid, are well-known substances of established structure

Phloroglucinol            Parahydroxybenzoic Acid
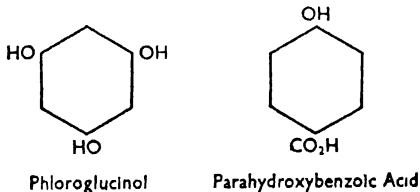
The structures represented are derived in a variety of ways and confirmed by numerous syntheses and inter-relations. For example, an application of Körner's principle may be made. The symmetry of the benzene ring demands the existence of three isomeric dihydroxybenzenes and three trihydroxybenzenes. In each case three isomerides and three only are known. Their transformations are as follows, and it will be appreciated that the constitutions assigned are the only ones in harmony with the facts represented by the arrows. In some cases the processes involve intermediate stages but some of the changes are direct.

Pyrocatechol    Resorcinol    Quinol

Pyrogallol    Hydroxyquinol    Phloroglucinol
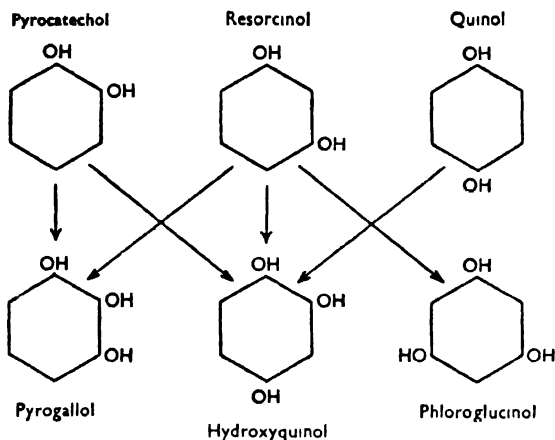
The *synthesis* of phloroglucinol from acetic acid has been achieved in several ways [One of which by way of ethyl sodio-malonate can be readily demonstrated. The product is recognised by taking advantage of the fact that a solution of phloroglucinol in aqueous alcoholic hydrochloric acid stains wood red], and all these confirm the structure based on theoretical arguments. The constitution of parahydroxybenzoic acid is related to that of quinol through the transformations :
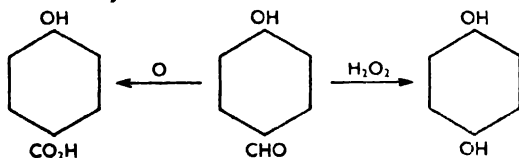


The formation of phloroglucinol and parahydroxy-benzoic acid from pelargonidin does not suffice to determine the constitution of pelargonidin chloride but taken into consideration with certain analogies a very plausible guess can be hazarded.

Pelargonin chloride and pelargonidin chloride

are obviously *oxonium* salts and among synthetic compounds no such stable salts, comparable to ammonium salts are known, unless it be those containing a specialised group known as *pyrylium*. This is a nucleus resembling a benzene ring as shown in the comparison :—
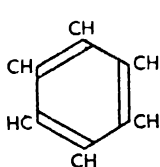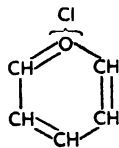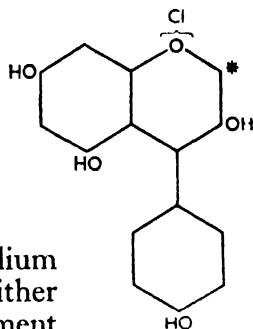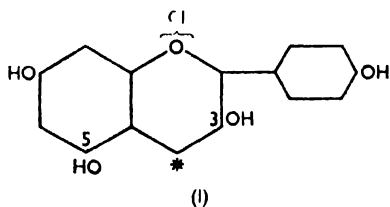


Benzine                Pyrylium Chloride

It was thus natural to assume that pelargonidin chloride is a pyrylium derivative and on this basis two structures suggested themselves



(I)

The HO group in the pyrylium nuclei cannot be situated in either benzene ring because replacement of HO by H never occurs in a potash fusion ; it cannot be in either of the positions indicated by an asterisk because HO so situated destroys the stability of pyrylium salts. This statement is based on the observed properties of synthetic compounds of known structure. Thus the HO must be in the position denoted 3, and hence the HO.$C_6H_4$. group cannot be in this position ; hence there are only two alternatives.

The arguments based on analytic studies take us no further, but (I) has been proved to be the correct representation of pelargonidin chloride by independent and unambiguous syntheses of that substance. We cover this point at a later stage by the synthesis of pelargonin itself.

Having now some conception of the nature of pelargonidin we may consider the manner of attachment of the sugar molecules. The evidence is connected with alterations of the properties of pelargonidin, or its close relatives, when the HO groups are modified or perhaps replaced by hydrogen. Many synthetic variations of pelargonidin were prepared and carefully compared with pelargonidin and pelargonin. Only two aspects of this work need be mentioned.

Pelargonin is fluorescent in alcoholic solution and on careful hydrolysis it is possible to obtain a monoglucoside of pelargonidin termed *pelargonenin chloride*, which shows a still brighter green fluorescence. [Demonstration of the formation of pelargonenin and its fluorescence in butyl alcoholic solution. The fluorescence is comparable with that of fluorescein and rhodamine, the constitutions of which are actually similar to that of pelargonenin in respect of the essential fluorophore. Pelargonenin does not occur in Nature ; it is isomeric with callistephin, the colouring matter of scarlet carnations and the 3-glucoside of pelargonidin.] Now no synthetic analogues of pelargonidin exhibit this property unless the HO in position 5 (I) is absent or replaced by RO. Hence we conclude that pelargonin and pelargonenin probably bear a $C_6H_{11}O_5.O.$ group in this position.

The second point is that the stability of pelar-

gonin in weak acid solution containing a trace of an iron salt is in marked contrast to the instability of pelargonidin and pelargonenin under the same conditions.

Again synthetic models indicate that this property is dependent on the circumstances at position 3. If HO is the substituent, the result is instability, but if RO or H are situated at this point the substances are markedly stable. [The contrast was demonstrated.]

Thus pelargonin should bear $C_6H_{11}O_5.O.$ at position 3 and on the basis of (I) for pelargonidin its constitution may be expressed by the formula:
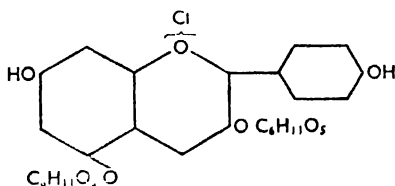


The reasoning was not conclusive but it led to the formulation of a good working hypothesis. Support was lent by further work which showed that the colour reactions of pelargonin in solutions of graded pH and under other conditions, were in harmony with the proposed formula, and indirectly by the proof that another plant pigment was constituted on the same general plan.

A successful synthesis of the pigment was required in order to remove the last doubt, and at this stage it is important to realise that the only connection between what follows and the preceding analytical phase is the above formula defining the object of the synthetical experiments.

It will have been gathered from the frequent reference to synthetic analogues of pelargonidin

that a generally applicable method of preparation of such pyrylium salts exists. This relies on the interaction of three components as shown in the scheme
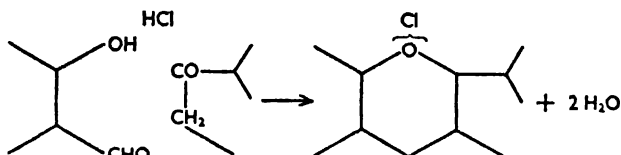


This reaction is so general, within certain limits found by experience, that it was possible to demonstrate the formation of a new pyrylium salt with its aid.

[2-Hydroxy-3-methoxybenzaldehyde (2.0 g.) and *w*-4-dihydroxy-3 : 5-dimethoxyacetophenone (2.7 g.) were dissolved in ethyl acetate (50 c.c.) and hydrogen chloride was passed into the solution for about 20 minutes. After 14 hours an excess of ether was added to the brown-red liquid precipitating the crude pyrylium salt as a brownish red microcrystalline powder. This was collected and extracted with boiling 2 per cent hydrochloric acid and a flocculent *picrate* thrown down by the addition of half saturated aqueous picric acid. The derivative was collected, dried and crystallised from aqueous alcoholic picric acid as hairfine dark brown needles which charred on heating and exhibited no definite melting point (Found in material dried at 100 deg. in a vacuum : C, 52.0 ; H, 3.3 ; N, 7.6. $C_{24}H_{19}O_{13}N_3$ requires C, 51.7 ; H, 3.4 ; N, 7.5 per cent). The related chloride was obtained from the picrate ; it dissolved in aqueous sodium carbonate to a violet red solution and in aqueous alcoholic sodium carbonate to a rich bluish violet solution. All the properties of

these substances were in agreement with those expected of 3 : 4'-*dihydroxy*-3' : 5' : 8-*trimethoxy-flavylium salts.*]

In our view the two components required for the synthesis of pelargonin chloride by this method are clearly
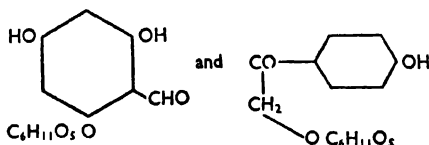


These have not been employed in quite their simplest form, but the modifications in each case are of an unimportant character and will be understood from the sequel. The phloroglucinol component (A) is made by way of the stages shown below.
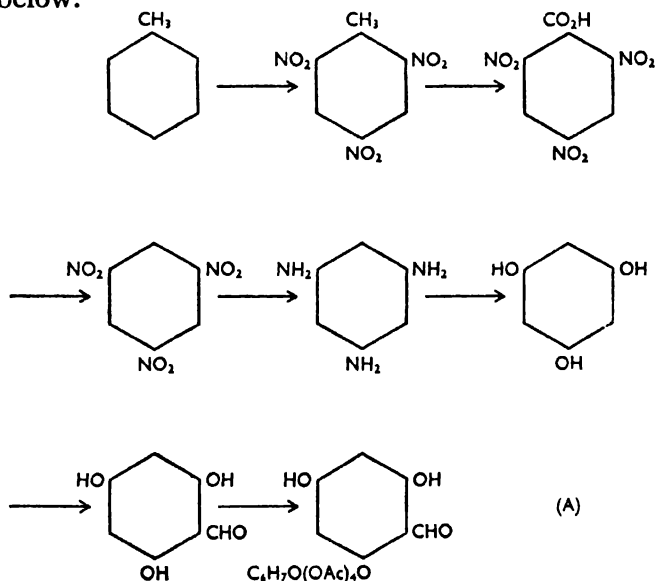
The formulae are self-explanatory to organic chemists and those not skilled in the art will not wish to examine the details. The glucose is introduced in the last stage in a protected form (acetylated) chiefly because the general synthetic reaction for glucosides makes use of tetra-acetylglucosidyl bromide in such stages as the following :

$C_6H_{12}O_6 = C_6H_7O(OH)_4OH \rightarrow$

$C_6H_7O(OAc)_4OAc \rightarrow C_6H_7O(OAc)_4Br$
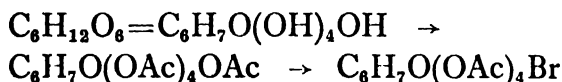
[The particular OAc replaced is that derived from the reducing group (see above).]

R.OH $\rightarrow$ RONa or ROAg and

RONa $+$ $C_6H_7O(OAc)_4Br$ $\rightarrow$ NaBr $+$ $C_6H_7O(OAc)_4OR.$
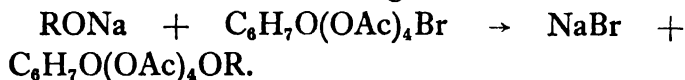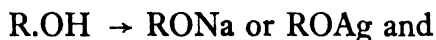
The action of alkalis under mild conditions changes the OAc groups to OH and we get a glucoside, $C_6H_{11}O_5.OR$. In our particular example, the substance to which the constitution (A) was assigned might have been (B).
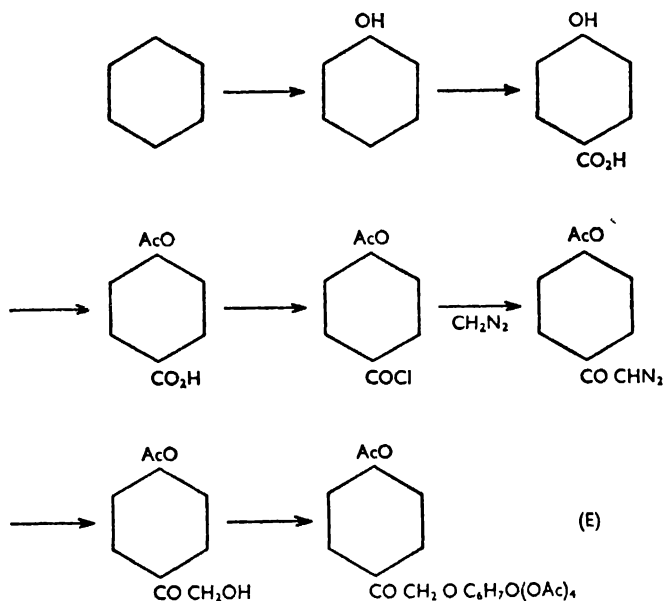


(B)    (C)    (D)
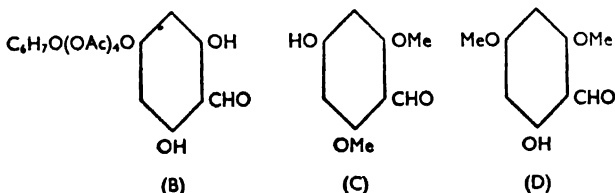
But when the two remaining OH groups of the glucoside were replaced by $OCH_3$ and the glucose removed, a substance was obtained which functioned well as a component in the general pyrylium salt synthesis mentioned above. Hence this pro-

duct must have OH and CHO in adjacent (ortho-) positions and must be (D) and not (C). Therefore the glucoside is (A) and not (B).

The second component (E) was also used as an *acetylated* glucoside; it could be made in a variety of ways, especially in the earlier stages. The following is one of these :

$$\text{C}_6\text{H}_6 \longrightarrow \text{C}_6\text{H}_5\text{OH} \longrightarrow \underset{\text{CO}_2\text{H}}{\overset{\text{OH}}{\bigcirc}}$$

$$\longrightarrow \underset{\text{CO}_2\text{H}}{\overset{\text{AcO}}{\bigcirc}} \longrightarrow \underset{\text{COCl}}{\overset{\text{AcO}}{\bigcirc}} \xrightarrow{\text{CH}_2\text{N}_2} \underset{\text{CO CHN}_2}{\overset{\text{AcO}}{\bigcirc}}$$

$$\longrightarrow \underset{\text{CO CH}_2\text{OH}}{\overset{\text{AcO}}{\bigcirc}} \longrightarrow \underset{\text{CO CH}_2\text{ O C}_6\text{H}_7\text{O(OAc)}_4}{\overset{\text{AcO}}{\bigcirc}} \qquad \text{(E)}$$

The stage is now set for the crucial experiment; (A) and (E) were together dissolved in ethyl acetate and hydrogen chloride introduced into the cold solution. The expected condensation occurred with formation of an acetylated diglucoside of pelargonidin chloride [one acetyl group was removed in the course of this process].
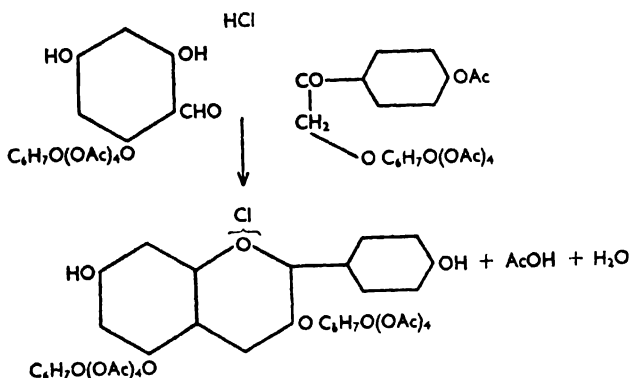
After solution in cold dilute alkali and regeneration of the salt by hydrochloric acid, pelargonidin 3 : 5-diglucoside chloride was obtained. After purification this salt proved to be identical in every respect with pelargonin chloride and thus the analytical reasoning received its final justification and the constitutional problem was solved.

The identity of the natural and synthetic pigments was proved by an exhaustive comparison of their properties and no divergences of any kind were noted ; the only difficulty was the complete purification of the natural colouring matter. Among the ·more characteristic properties, the indicator range [colours in solutions of graded pH], the distribution ratios between immiscible solvents, the form and habit of the crystals, the rate of decolorisation under appropriate conditions, the optical rotatory power [as yet unpublished], the absorption spectrum and the behaviour with solvents and reagents were carefully examined.

Although the technical details of this narrative may, and must, elude those who have not studied the science of organic chemistry, it is hoped that some impression has been given of converging

lines of attack on the problem of molecular structure. Only the purely chemical aspect has been touched upon and only the classical methods of our science have been pressed into service. Nowadays our resources have been greatly enlarged by the developments of physics and, especially in this field, by the X-ray analysis of crystals. Nevertheless these newer complementary weapons owe much of their sharpness to the existence of the system of the carbon compounds, that system which Sir Frederick Hopkins has truly described as one of the greatest achievements of the human mind. The verification of prediction, provided it be a long shot, is a most satisfying experience and can never become a commonplace.

[R. R.]

## SENSITISING DYES AND THEIR APPLICATION TO SCIENTIFIC PHOTOGRAPHY

ONE of the problems to be encountered in the early days of photography was that photographic materials were not sensitive to the spectrum in the same way that the eye is. Whereas the normal eye is sensitive to the colours of the spectrum from violet through blue, green, yellow and orange to red, the early photographic plates were sensitive only to the ultra-violet, violet and blue.

In 1873 H. Vogel, of Berlin, found that some collodion dry plates possessed some sensitivity in the green portion of the spectrum, and he traced this green sensitivity to the presence in the emulsion of a yellow-red dye. Vogel got similar effects with other dyes which, on this account, were later called sensitising dyes or sensitisers, and he came to the conclusion that when a dye acted as a sensitiser in this way, the region of the spectrum for which it was effective corresponded approximately to the absorption band of the dye. Thus a dye which absorbed green light would render a photographic plate sensitive to green. This fundamental relationship underlies all work on sensitising, and it is worthy of attention that Vogel

grasped this truth immediately in spite of the fact that his emulsions were very slow, his dyes probably impure and, at best, weak sensitisers, and his apparatus primitive.

Vogel recognised that the dyes which acted as sensitisers for silver bromide must not merely stain the collodion or gelatin in which the silver bromide was imbedded, but must actually be adsorbed on to the surface of the silver bromide particles. Thus the energy which they absorbed could be transmitted in some way to the silver bromide.

Vogel's results were at first subjected to a great deal of adverse criticism, and even ridicule, many other workers being unable to confirm his results, but Becquerel in 1874, and then Waterhouse in 1875, were able to support his discovery.

The first sensitising dye to attain any general use was eosin, which sensitises collodion emulsion strongly for the yellow-green. After the introduction of gelatin dry plates, J. M. Eder suggested the use of erythrosin in the place of eosin, and this was generally adopted for the production of the so-called 'orthochromatic' plates, and was almost the only dye used in commercial manufacture before 1906.

Shortly after the beginning of this century attention was directed to the dyes belonging to the 'cyanine' group. In 1856 a dye known as 'cyanine' had been made by Williams from impure quinoline, but although it had a beautiful blue colour, it was very fugitive to light. Williams' cyanine had been used by Vogel in his early experiments and was considered a useful sensitiser for the orange-red, but it caused fog and spotty patches on the plates.

In 1883 another dye made from quinoline and related to cyanine was discovered, and in 1902 this dye, known as 'ethyl red,' was tested by Miethe and Traube for sensitising power, and was found to give good green sensitivity without the tendency to fog characteristic of cyanine. Ethyl red belongs to a group of dyes known as the isocyanines, which are synthesised by the treatment of a mixture of quinaldine and quinoline quaternary salts with an alkali.

At the time that Miethe and Traube described the sensitising power of the simple isocyanine, ethyl red, E. König of the Hoechst Works of Meister, Lucius and Brüning, was synthesising a number of isocyanines, which were placed on the market under trade names, such as Pinachrome, Orthochrome T, Pinaverdol, etc., these being made from substituted quinolines and quinaldines, and by 1904 it was clear that the cyanine dyes of the new type were much more suitable for use in photographic sensitising, especially for the yellow-green and orange, than any dyes previously known.

In 1904 Homolka, one of König's colleagues, carried out the reaction used in the preparation of an isocyanine but with the addition of formaldehyde, and in the place of the red isocyanine he obtained a blue dye, of which the structure was not elucidated at the time but which proved to be a powerful sensitiser for the red, and was placed on the market under the name of 'pinacyanol.' It was later found that the quinoline nucleus does not form part of the dye molecule, and pinacyanol is at present made from the quinaldine salt, with a compound such as iodoform or orthoformic ester to supply the additional methine group.

Soon after the discovery of pinacyanol the

little firm of Wratten and Wainwright in Croydon applied the new dyes, especially pinachrome and pinacyanol, to the preparation of commercial plates, and in 1906 made plates containing both pinachrome and pinacyanol, which were placed on the market under the name of 'Wratten Panchromatic' plates. These were the first panchromatic plates to be made on a commercial scale with the new type of dyes, and following the introduction of the plates and of a series of light filters adapted for use with them, the whole art of the photography of coloured objects developed rapidly as a practical commercial process.
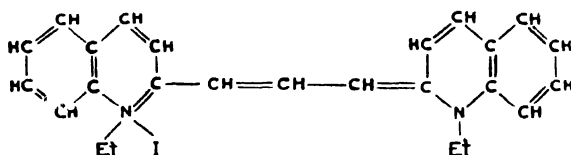


PINACYANOL ETHIODIDE

FIG. 1.

The cyanine dyes are derived from organic bases containing nitrogen. Bases of this kind add on two atoms or groups to form quaternary salts. Thus quinoline adds on ethyl iodide and forms quinoline ethiodide. The bases from which the first cyanine dyes were derived were quinoline, and quinaldine, in which the hydrogen attached to the carbon atom of quinoline next to the nitrogen has been replaced by a methyl group.

Our present knowledge of the structure of the cyanine dyes is due very largely to the work of Dr. W. H. Mills and his students at Cambridge, and of especial importance was the paper that he published with Dr. F. M. Hamer in 1920, on the structure of pinacyanol, which is shown in Fig. 1.

It will be seen that the two nitrogen atoms in the quinoline nuclei are connected by a chain of five carbon atoms with alternate double and single bonds between them. A system of this kind is known as a conjugate chain and is characteristic of the cyanine dye series.

Since the dyes having one CH group, derived from the base itself, uniting the two nuclei are called cyanines, Mills and Pope called the dyes with three CH's in the chain 'carbocyanines,' because it was necessary to supply an additional carbon atom. Later Heilbron and his students at Liverpool made dyes with five CH's in the
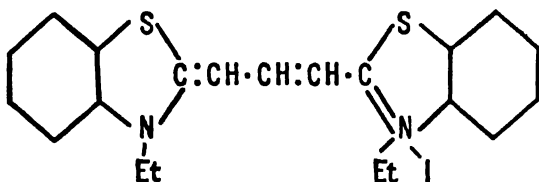


THIACARBOCYANINE

FIG 2

chain and these he called 'dicarbocyanines.' Very important dyes have been made containing seven CH's which-are known as 'tricarbocyanines,' while, as will be seen later, tetra- and penta-carbocyanines, have been prepared, with chains having nine and eleven CH's respectively.

It will be realised that there must be an enormous number of cyanine dyes, because we not only have the possibility of varying the number of CH's in the chain but we can substitute a hydrogen in the chain by another group, and we can also prepare dyes from a great many different organic bases containing nitrogen.

More than 50 years ago Hofmann made a

violet dye from the quaternary salt of 1-methyl benzthiazole and Hofmann's dye was shown by Mills to be a thiacarbocyanine, the exact analogue of pinacyanol. Its structure is shown in Fig. 2.

A very large number of dyes has now been made from substituted benzthiazoles, and many of these are excellent sensitisers.

The development of these new carbocyanines has during the last few years been of the greatest value for practical photography. The new dyes from the thiazole bases are far better sensitisers than pinacyanol and the other dyes derived from quinoline, and by making use of them new pan-chromatic materials have been prepared which are generally known as 'super-sensitive' or 'hyper-sensitive' materials.

The great value of these new sensitisers is that they do not lower the general sensitivity of the emulsion, and that they show no tendency to fog, emulsions sensitised with them often being even less prone to fog than the unsensitised emul-sions. As a result, the effective sensitivity of the new panchromatic materials is greater than that of the same emulsions unsensitised, and this is particularly true when the emulsion is of very fine grain and is consequently slow before sensi-tising, and where the exposures are made to a light source rich in the red and green rays of the spec-trum. Just at the time that the new sensitisers became available, the motion picture industry adopted the use of filament lamps for the greater part of studio lighting, in place of the arc lamps used previously. The filament lamps are more convenient in operation because they need much less attention and they have the advantage of being entirely silent, which was necessary when sound

recording was added to the studio problems. But, nevertheless, it is doubtful whether the filament lamps could have been used had it not been for the great increase in the sensitivity to yellow light which was available from the application of the new sensitisers. The very great change which has occurred in the methods used in the making of motion pictures is, therefore, directly traceable to the scientific work on the structure of the cyanine dyes which I have been discussing. In the same way, the use of panchromatic materials, and especially of panchromatic materials of exceptionally fine grain, stimulated the use of miniature cameras, and notably the use of those miniature cameras for photography indoors by artificial light which is often referred to as the use of the candid camera, perhaps because the results are sometimes more candid than pleasing.

The introduction of the new panchromatic materials, therefore, marks a real event in the history of photography and I think that it is probable that the years 1928 to 1930 may be regarded as beginning a new period in photography, dating from the introduction of panchromatic materials into everyday use, as distinguished from their previous employment for the specific photography of coloured objects. For the scientific applications of photography, however, the increase of sensitivity obtained in the visible spectrum, while of value, was less significant. The effective sensitiveness of the new panchromatic materials as compared with those available previously may be three or four times, and this is of the greatest importance in motion picture work, but it is not sufficient to make any great change in the practice of astronomical

or spectroscopic photography.   A much more im-
portant matter was the development of sensitisers
specifically adapted for use in the infra-red region
of the spectrum.

The first sensitiser to give satisfactory results
in the infra-red was dicyanine, which was made
at Hoechst in 1906.  This dye was a true sensitiser
for the near infra-red, but it was very unsatisfactory
in use, and although it enabled the photography of
the infra-red spectrum to be carried out, it was
only used by a few workers who had acquired the
necessary skill, and it was not practicable to make
ready-sensitised plates with it.  In 1919 Adams and
Haller, working at the United States Bureau of
Chemistry, carried out the pinacyanol reaction,
but instead of using quinaldine, in which the
methyi group is on the carbon next to the nitrogen,
they used lepidine, which is identical with quinal-
dine except that the methyl group is further re-
moved from the nitrogen, being separated by
another two carbon atoms.   The synthesis occurs
quite normally and yields a blue-green dye which
Adams and Haller called kryptocyanine.

The structure of kryptocyanine is shown below:
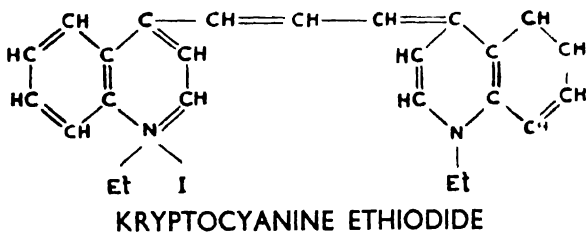


KRYPTOCYANINE ETHIODIDE

FIG. 3.

It will be seen that it is a carbocyanine like
pinacyanol, but that in addition to the three
CH's of the chain the nitrogen atoms are separated

by three carbon atoms in each nucleus, so that the conjugate chain of kryptocyanine has four more links than that of pinacyanol. As we should expect, this shifts the absorption of kryptocyanine, and also its sensitising, towards the region of longer wavelength, so that its maximum of sensitivity is in the near infra-red at λ7500A.

Kryptocyanine is a very strong sensitiser and by its use it was at once possible to take ordinary photographs using infra-red light. As long ago as 1910, Prof. R. W. Wood took photographs by light of longer wavelength than λ7000A, using a strong filter to cut out all the shorter wavelengths, and plates sensitised with pinacyanol, and in these photographs the notable peculiarities of landscape pictures taken by infra-red light were manifest. The blue sky is almost black owing to the low scattering power of the upper atmosphere for the longer wavelengths, clouds stand out in startling contrast to the black sky, and the chlorophyll of grass and trees reflects the infra-red so that foliage appears as if it was covered with snow. Moreover, because of the low scattering power of the atmosphere for the longer waves, pictures taken by infra-red light show a penetration of the distance greatly exceeding that shown by ordinary photographs or even that visible to the eye. For Wood's pictures, the average time of exposure was approximately five minutes, but with kryptocyanine, an exposure of a fraction of a second was sufficient.

When preparing kryptocyanine in our laboratory in Rochester, H. T. Clarke found that the solution contained another dye which we named 'neocyanine', and which we now know to be of rather complicated structure, although it is still a cyanine

dye. With this dye the maximum sensitivity was at $\lambda 8500A$ and by means of it the photography of the spectrum was easily extended as far as $\lambda 9000A$.

About 1930 the tricarbocyanine dyes were synthesised almost simultaneously by three groups of workers, and their very powerful sensitising and ready accessibility made it possible to sensitise throughout the infra-red spectrum as far as $\lambda 11000A$. Of the tricarbocyanines, the one having the furthest extension in the infra-red was xeno-cyanine, which is the tricarbocyanine from lepi-dine corresponding exactly to kryptocyanine, except that there are seven CH's in the chain and no less than thirteen CH's between the two nitrogen atoms, seven of them being in the chain and three in each of the nuclei. The maximum sensitising of xenocyanine is at $\lambda 9600A$.

The structure of xenocyanine is shown below :
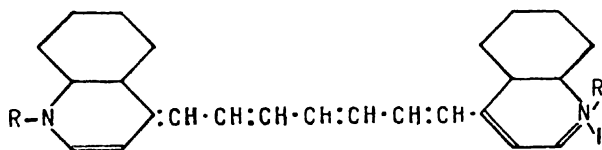


XENOCYANINE

FIG 4.

Using tricarbocyanines a great many very remark-able photographs have been taken by infra-red light. From air-planes photographs can be taken which far exceed the range of the human eye, the record for distance still being held by that taken by Capt. A. W. Stevens in 1932 of Mount Shasta in California, the plane at the time flying over the Pacific 330 miles from Mount Shasta.

In addition to these long distance photographs, of which many examples have been made during

the last few years, photographs may be taken by infra-red light which is effectively invisible to the eye.   As long ago as 1931 a photograph was taken of a large group of people who could see nothing whatever at the time, the room being lighted by tungsten lamps covered by filters of so great a depth that there was no visible light at all in the room.

Within the last year a still further extension of photography into the infra-red has become possible as a result of the development of tetra- and pentacarbocyanines.   The method by which these long-chain dyes are prepared is due largely to Dr. W. König of Dresden.   In this synthesis a compound of what is known as a 'dianilide' the chain of carbon atoms is supplied by taking type and allowing this to react with the quaternary salt of the base.   Aniline is split out and the dye is formed.   By an extension of this system tetra- and penta-carbocyanines have been prepared, of which the latter has no less than 11 carbon atoms in the chain, and these dyes have proved to be excellent sensitisers for the far infra-red. With them Dr. Meggers of the Bureau of Standards has recorded an argon line at $\lambda 13008A$ and Dr. Babcock at Mount Wilson has photographed the solar spectrum to between $\lambda 13000$ and $14000A$.

There is thus a very wide range of cyanine dyes in which the absorption band and sensitising region ranges from extreme violet, or even ultra-violet in the case of such dyes as the oxacyanines, to a point in the infra-red approaching that where water vapour begins to absorb ; this makes it possible to prepare photographic plates of high sensitivity to any special region in the spectrum,

VISUAL AND PHOTOGRAPHIC SPECTRA

VISIBLE

PHOTOGRAPHIC THRU GLASS BEFORE 1875

PHOTOGRAPHIC WITH QUARTZ BEFORE 1904

PHOTOGRAPHIC 1904-1919

PHOTOGRAPHIC 1919-1925

PHOTOGRAPHIC 1925-1931

PHOTOGRAPHIC 1932-1934

PHOTOGRAPHIC 1934 -

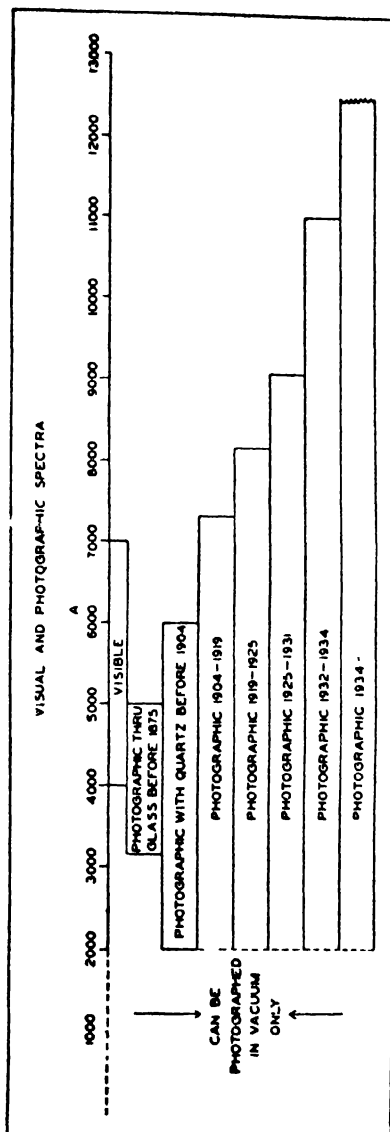CAN BE PHOTOGRAPHED IN VACUUM ONLY

FIG. 5.

and such plates are, naturally, of the greatest interest to scientific workers, and especially to astronomers and spectroscopists.

A few years ago I classified the plates required for these purposes, in order to reduce the mass of material to a system which would be available for the scientific worker.   In the first place there are various types of photographic emulsions which are useful in scientific work and we must have some method of indicating these.   The fastest type of emulsion was called type 1, this having the disadvantage of a somewhat grainy structure and being of only moderate contrast.   Type 2 is a somewhat slower emulsion of greater contrast and less graininess, while type 3 is a fine-grained emulsion of high contrast.   Type 4 was of extreme contrast and still less speed, and type 5 a very slow emulsion of extremely fine grain.   Now these different types of emulsion can be sensitised for different parts of the visible and infra-red spectrum, and the different sensitisings are indicated by means of letters, so that it is possible for an astronomer, for instance, to decide that he could afford the exposure necessary for a type 3 emulsion and that he wanted his maximum sensitivity in the orange, with as high sensitivity as possible through the green but with no particular sensitivity in the red.   Such a result is shown with the sensitising known as D, and the particular plate that he required would be classified as 3D.

The progress which the development of dyes has enabled us to make in the photography of the spectrum may be summed up in an approximate form in the chart shown in Fig. 5.

At the top of the chart is shown the visible spectrum and that portion of it which could be

photographed on a collodion plate through glass apparatus before Vogel discovered the use of sensitising dyes. Vogel's discovery extended the photographic spectrum to λ6000A. The discovery of pinacyanol in 1904 extended the photographic spectrum to beyond λ7000A. The effect of dicyanine has been ignored in the construction of the chart, though for a few workers its use extended the photographic spectrum to beyond λ8000A. It was, however, only easy to reach that wavelength after the discovery of kryptocyanine in 1919, while the production of neocyanine in 1925 extended the range to λ9000A. Xenocyanine, made first in 1932, enabled λ11000A. to be reached, and the penta-carbocyanine has now extended the photographic spectrum to beyond λ12000A.

[C. E. K. M.]

# Friday, March 6, 1936

JOHN S. HIGHFIELD, M.I.E.E.
*Manager and Vice-President, in the Chair*

SIR WILLIAM BRAGG,
O.M., K.B.E., M.A., D.Sc., F.R.S., M.R.I.
*Fullerian Professor of Chemistry*

## THE ELECTRIC PROPERTIES OF CRYSTALS (I)

THE origin of the magnetic properties of substances has been, and still is a problem of great difficulty. It cannot conveniently be expressed in terms of the phenomena observed on a larger scale. Some substances are diamagnetic. Their presence in a magnetic field, in which the magnetic induction is maintained at a constant value in all parts except in the neighbourhood of the substances, implies an increase of energy, so that if they are placed in a field which varies in strength from place to place they are urged to move from the stronger to the weaker parts of the field. On the other hand, paramagnetic substances are drawn towards the stronger parts. Ferromagnetic substances are distinguished by excessive paramagnetism and also by the display of hysteresis and other complicated effects.

Long ago Weber sought to explain diamagnetism as due to the presence in the molecule of channels in which electric currents could circulate without encountering resistance. The imposition of a magnetic field would be resisted by currents

circulating in those channels; and the magnitudes of the currents would necessarily be, according to classical theory, just sufficient to prevent any lines of induction from threading the circuits in which they ran. The lines would be forced to go out of their way, and so additional strain would be placed on the medium. The energy of the system would be increased. Clerk Maxwell discussed Weber's Theory in his " Electricity and Magnetism."

When the existence of the electron was realised, and it became clear that many atomic properties could be explained as consequences of orbital motions of electrons about a nucleus, the possibility that magnetism in its various forms might be included among those properties came naturally into view. In 1905 Langevin published his famous dissertation on the subject in the Annales de Chimie et 'de Physique.* The electron orbits in the molecule were supposed to be fixed with respect to the molecule and to each other. If the magnetic moments of the orbits balanced each other, the molecule as a whole had no magnetic moment, and the molecule was diamagnetic ; but if there was lack of balance and the molecule could be forced therefore to alter its orientation, the paramagnetism far exceeded the diamagnetism and reversed the sign of the moment which the molecule assumed under the action of the field.

Langevin's theory gave a satisfactory explanation of many of the phenomena of magnetism. For instance, the thermal movements of the molecules of a gas should have no effect on the diamagnetism, since they would not disturb the general isotropy. On the other hand, when

* Ann. Chim. et Phys., 8.S. Vol. v, p. 245. 1905.

paramagnetic molecules were urged by the imposed field to a common orientation, the thermal movements would tend to break up the arrangements. Both these deductions from the theory agree with experiment. The paramagnetic susceptibility varies inversely as the absolute temperature, broadly speaking.

The modern development of wave mechanics provides modes of expression and argument which, in a formal way, give a fuller explanation of the phenomena. Nevertheless they lead to deductions which do not differ greatly from those of the older theory, so that the simpler terms and conceptions of the latter may still be used, at least when the main facts are considered. This is done in what follows.

One of the main difficulties in the way of tracing diamagnetism to its source has been the fact that almost all the experimental measurements give average values for all possible orientations of the molecule. The substance has been in the form of a gas or a liquid or an amorphous solid or a crystalline solid consisting of an agglomerate of crystals pointing in all directions. Now the molecule is generally anisotropic, sometimes highly so. If the average of susceptibilities in all directions is all that is found, a very important characteristic of the molecule is ignored, and it is not to be expected that progress will be made.

The situation is changed entirely when the positions of the molecules in the unit cell are known. It is then possible to deduce the susceptibilities of the single molecule from those of the crystal, which latter can be measured. It is true that the quantities thus determined are related to the molecule as it is built into the

crystal. But it appears that the influence of neighbouring molecules on each other's suscepti-bilities is quite small. Thus molecular quantities are determined which are variants of the mole-cule, a result of great importance.

Thanks to the work of Dr. J. M. Robertson* and his colleagues in the laboratories of the Royal Institution, and to other workers, the structures of a number of crystals have been determined with great accuracy, and the results have been used for the calculation of the constants of the corresponding molecules.

The connection between the constants of the molecule and those of the unit cell may be used in another way. If the latter have been measured and the former can be assumed with a sufficient probability, useful conclusions may often be drawn as to the positions of the molecules in the unit cell. This converse treatment does not lead to definite results because the constants of the molecule are never known with exactness *a priori*. They can, however, be estimated in many cases, particularly in that of certain organic molecules.

In the hands of Krishnan† and his fellow workers this method of contributing to the know-ledge of crystal structure has been used with very good results. In the course of their work valuable determinations of the susceptibilities of a large
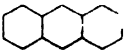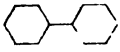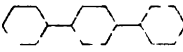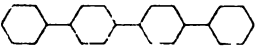
* " The Crystalline Structure of Naphthalene." Roy. Soc. Proc. A. Vol. cxlii, p. 674. 1933. "The Structure of Benzoquinone." Roy. Soc. Proc. A. Vol. cl, p. 106. 1935. " X-ray Analysis of the Dibenzyl Series." Prasad & Woodward. Roy. Soc. Proc. A. Vol. cliv, p. 197. 1936.

† Investigations on Magne-crystallic action. Part I. Diamag-netics. Roy. Soc. Phil. Trans. A. Vol. ccxxxi, p. 235. 1933. Part II. Paramagnetics. Roy. Soc. Phil. Trans. A. Vol. ccxxxii, p. 99. 1934. Part III. Further Studies on Organic Crystals. Roy. Soc. Phil. Trans. A. Vol. ccxxxiv, p. 265. 1935.

number of crystals have been made. These authors were the first to make good use of this method of calculating the positions of the molecules in the unit cell.

## TABLE I

### DIAMAGNETIC SUSCEPTIBILITIES OF CERTAIN ORGANIC MOLECULES.

| | $-K_1 \times 10^{-6}$ | $-K_2 \times 10^{-6}$ | $-K_3 \times 10^{-6}$ |
|---|---|---|---|
| Naphthalene | 56·1 | 53·9 | 169 |
| Anthracene | 75·8 | 62·6 | 251·8 |
| Biphenyl | 66·9 | 66·9 | 174·7 |
| p-diphenyl-benzene | 96·8 | 88·1 | 271·3 |
| p-diphenyl-biphenyl | 122 | 110 | 372 |

$$\longleftarrow K_1$$
$$K_2$$

Mrs. Lonsdale and Professor Krishnan have recently contributed to the Proceedings of the Royal Society* a paper in which they have discussed the methods of deriving the susceptibilities of the molecule from those of the crystal in various cases. A few of their results are given in Table I, in order to show their remarkable character. The molecules in question are all flat, and in fact their shapes are well indicated by the formal representations of the table. The three principal axes may therefore be assigned without

* In the press.

hesitation to the length of the molecule, to its breadth and lastly to the normal to its plane. The susceptibilities of the gram-molecule in these three directions are denoted in the table by $K_1$, $K_2$ and $K_3$, and the figures given are multiples of these quantities by $-10^6$.

The first striking feature is the anisotropy. If the diamagnetism is ascribed to the existence of electron orbits, as in Langevin's theory, it is clear that the planes of the orbits are strongly oriented with respect to the molecule and on the whole are disposed to be parallel to it. They must be bound to the molecule by forces sufficient to preserve the orientation.

There are also remarkable additivities. The comparisons between the first three molecules on the list may be extended to take in benzene, for which the constants may be calculated indirectly to be nearly 37, 37 and 90.

It is to be observed also that the existence of these additivities, as well as the constancy of the contributions to the susceptibilities furnished by atoms and combinations of atoms, shows that the diamagnetism induced in any molecule has very little influence on that of its neighbours.

This short description of a new and most interesting development may serve as an introduction to the papers in which it is more fully discussed. References to some of the more important of these papers are given above.

The discourse was illustrated by demonstrations showing the magnetic properties of substances, especially the diamagnetism of crystals and the anisotropy of the molecule.

[W. H. B.]

SIR ROBERT ROBERTSON, K.B.E., M.A., F.R.S.
*Treasurer and Vice-President, in the Chair*

DR. H. FREUNDLICH
*The Sir William Ramsay Laboratories of Inorganic and
Physical Chemistry, University College, London*

## STRUCTURES AND FORCES IN COLLOIDAL SYSTEMS

WE use the term ' colloidal ' in a somewhat different way from Thomas Graham who introduced it seventy-five years ago. Graham called colloids two kinds of systems : colloidal solutions or sols, like those of gelatine, proteins, metal oxides, etc., and gels, like the jelly of gelatine or those of the juice of fruits, i.e., jellies of pectine, etc. He believed them to be distinguished by two special properties ; the particles of colloidal solutions are larger than those of normal solutions, e.g., of sugar or salt in water ; they are not simple molecules with a diameter of about one-hundred millionth of a centimetre, but are ten to ten thousand times larger. That is why they do not dialyze, i.e., do not pass through the pores of membranes like parchment. This first property is one of size, of degree of dispersion. Graham assumed furthermore that the particles of colloidal systems were amorphous, because on evaporating a colloidal solution, or on drying a gel, no large crystals are formed but amorphous masses like the globules of starch or a dry sheet of gelatine. This second criterion concerning the state of aggregation did not stand the test

of time. We have a much more powerful method for distinguishing amorphous and crystalline structure : X-ray analysis. When applied to colloidal systems it was found that in very many cases their particles are micro-crystalline. It would have been artificial to call only those systems colloidal which had amorphous particles ; the difference in behaviour between such systems and those having crystalline particles is not sufficiently great. So one gave up the second criterion and confined oneself to the first : colloidal systems are only distinguished by their degree of dispersion ; they are two-phase systems with particles too small to be seen in a microscope, but larger than normal molecules. Systems containing very large molecules, like the sols and gels of proteins, have colloidal properties. Thus it is no contradiction if one speaks of ' crystals ' in colloidal systems, or shows that the latter may manifest properties generally found only in crystalline systems.

In dilute colloidal solutions we need not deal with special structures, as a rule the particles are distributed at random, travelling about in lively Brownian movement. In concentrated sols structures are readily found. They are the cause of such sols being easily transformed into gels. It is a well-known fact that this transformation can be caused by change of temperature ; if a sufficiently concentrated aqueous solution of gelatine is prepared at higher temperatures it turns to a jelly when cooled down ; on heating it is liquefied again to a sol and this transformation may be repeated practically indefinitely. Sols and gels are so closely related that they may be transformed into each other even without any

change of temperature, simply by shaking or stirring at constant temperature. This phenomenon has been observed occasionally for quite a long time, but it was not realised until a short time ago that it is most frequent and important in biological and technical processes. Instead of the cumbrous term 'isothermal reversible sol-gel-transformation' the term 'thixotropy' and the adjective 'thixotropic' have been introduced. A concentrated sol of iron oxide such as investigated by Graham may easily be made thixotropic by adding a small amount of an electrolyte like NaCl ; the jelly formed is liquefied by shaking and the sol sets again to a jelly in a well-defined time. This process may be repeated indefinitely. The change is going on throughout the bulk of the colloid, not only on its surface, as was supposed by some investigators. Experiments with a transparent jelly of $Al_2O_3$ prove this as certain ; air bubbles or small glass balls may be seen to come to a standstill in the interior of the colloid when the liquid sol solidifies to a jelly.

Thixotropy has not only been observed with colloidal systems in the strict sense of the word, containing exclusively extremely small particles not visible in the microscope. It is found very frequently in coarser suspensions of clays, or of powders of many minerals such as slates, etc. The minerals must be powdered sufficiently finely ; a large percentage of the particles must have a diameter of about 1 $\mu$, that is, a tenthousandth of a centimetre. An example of this kind is finely powdered Solnhofen slate mixed with water. In this special case the transformation proceeds so rapidly that the liquefaction

may be heard when the tube is shaken close to the ear, but it cannot be seen because solidification proceeds too quickly.

It is surprising that a fairly large amount of liquid, sufficient to allow the particles to move in lively Brownian movement, may also be entirely enclosed when the system sets to a gel. This is especially surprising when we consider the limiting cases ; the sol of $V_2O_5$, for instance, whose particles are markedly rod-shaped, may form thixotropic gels whose concentration is only ·1 per cent, i.e., we have more than 99 per cent of an aqueous solution. In cases like this it is perhaps most probable that the gel is traversed by a fibrillar structure containing the liquid in its meshes. But the ease with which the structure is destroyed by mechanical forces, and is built up again spontaneously in a well-defined time, seems to prove that the particles of the fibres are not bound to each other by primary valencies, as they are in normal chemical compounds, but by weaker forces. One is dealing with a phenomenon very similar to the one called "coagulation of a sol," which takes place if we change the concentration of the sol and the amount and nature of the electrolyte added. If this is done, for instance, with a concentrated $V_2O_5$ sol, instead of a thixotropic gel, a thick fibrous coagulate of $V_2O_5$ is formed, which cannot be liquefied again by shaking. The action of the electrolyte is the same in both cases. In the original stable sol the particles do not stick to each other, they may approach each other ever so closely, yet repelling forces will separate them from each other again. It is fairly certain that this repelling force is caused by the electrical charge of the

particles. The distribution of ions on and round the particles, is such that this repulsion acts strongly whenever the particles come close to one another. But when the concentration of ions is changed by adding another electrolyte, repulsion is reduced, the particles stick together as soon as they approach each other sufficiently closely and thus cause in small concentrations thixotropy, in higher concentrations coagulation. Forces of attraction make themselves conspicuous, whose nature is still a matter of controversy. They seem to be more far-reaching than those molecular forces assumed by van der Waals ; but modern theory would perhaps allow for an effect of this kind, large enough to be acting at an appreciable distance from the particles. Anyhow, in coagulation the particles approach each other very closely, and then enclose but a small amount of liquid ; in thixotropy they remain further apart, or perhaps only a fraction participates in forming a structure. Here a very much larger amount of liquid is enclosed. Thus thixotropy is, in many cases, a preliminary stage of a mild, reversible kind of coagulation.

The other limiting cases are the pastes made up of more coarsely powdered minerals. The amount of solid substance necessary for producing a thixotropic mass is much larger, namely, from 10 to 70 per cent, and it is, therefore, not so surprising that these suspensions have the properties of solids. Here the correlation with the question of packing is especially interesting. In the pastes we are dealing with a large number of particles are subjected to forces, like those of molecular attraction or gravity, which tend to pack the particles more or less closely. If these

particles are all of the same size and of well-defined shape, for instance, spheres, there exists a limiting number of grades of packing ; packing may be extremely close, one sphere touching twelve neighbouring spheres, or it may be loose, one sphere touching only four neighbours ; there are intermediate cases, one sphere touching ten, eight or six neighbours. The volume of the interstices not filled by spheres grows when the packing becomes looser: the relative volume changes from ·26 for the case of close packing to ·66 in the case of loose packing. It is not possible to apply these considerations to thixotropic systems in a strict form. The particles are not sufficiently equal in size and well-defined in shape, but very probably the main point still holds ; the volume of the free space grows when the packing becomes looser. Several facts which were observed may be explained from this point of view.

Thixotropy implies that the solid system encloses a fairly large amount of liquid, or else the particles would not have space enough for their translations in Brownian movement when the system is liquid. Evidently a loose packing ought to be advantageous for producing a thixotropic behaviour. As an indicator of the degree of packing the sedimentation volume of the particles may be used. This can be determined by mixing more liquid with the powder than is necessary to produce a thixotropic paste, and to measure the height of the volume of powder after it has settled in a graduated tube. It is, indeed, a fairly frequent rule that a powder having a large sedimentation volume in the liquid is able to give thixotropic pastes with it, whereas

if the sedimentation volume is small, the mixture of particles and liquid is not thixotropic.

Thus thixotropy seems to be the counterpart of another phenomenon observed with mixtures of powders with liquids and bound to a close packing ; this is the phenomenon, called dilatancy, discovered and discussed 50 years ago by Osborne Reynolds. It is frequently observed on the moist sand of a beach : when one treads upon it, it becomes hard and dry, but it turns moist again as soon as the foot is lifted. This may be explained in the following way : sand, mixed with water, tends to assume a very close packing ; according to Spring, a fine sand, mixed with water, encloses, when left to itself, 26 per cent of its volume in water, i.e., it is the closest packing possible. The sedimentation volume is correspondingly small. If external forces displace the particles, the degree of packing becomes less close, the volume of the interstices increases and the amount of liquid present is not sufficient to fill them, consequently the sand turns hard and dry ; when the external forces stop acting, the particles will return to their original close packing, the interstices are being filled up, and the sand becomes moist again. Osborne Reynolds showed this behaviour in a lecture experiment by filling moist sand into a rubber balloon which communicated with a tube filled with an excess of liquid. If the particles of sand are displaced by external pressure the liquid is not pressed out as might be expected, but is sucked in because the volume of the interstices increases.

There is still a further property of these powders, which is correlated with their degree of packing and which may be briefly mentioned.

It is a phenomenon which we are used to call 'plasticity'; when we say, for instance, that a piece of moist clay is very plastic. Obviously, a mass showing dilatancy, a mass which becomes hard when we try to deform it, is not plastic. Thus plasticity, like thixotropy, is linked up with a certain loose packing. Therefore, the following correlation of properties holds fairly well :

TABLE I.

| CLOSE PACKING | LOOSE PACKING |
|---|---|
| Small volume of sedimentation | Large volume of sedimentation |
| dilatancy | no dilatancy |
| no plasticity | plasticity |
| no thixotropy | thixotropy |

Concentrated suspensions of clays in water and aqueous solutions are, therefore, strongly thixotropic.

Pure moist quartz sand is naturally closely packed and not thixotropic, but shows marked dilatancy ; yet it may be readily transformed into a loosely packed and thixotropic system, provided that, to start with, the particles are sufficiently fine, a condition which is not essential for dilatancy. One only need add a few per cent of a fine clay which is strongly ' hydrophilic,' i.e., which has a marked affinity to water ; also, other hydrophilic colloids like lecithine, etc., may be used. These substances probably coat the quartz particles with a thin layer having a strong affinity to water, and thus causing a fairly thick layer of liquid round the particles—the particles are thus not more closely packed.

The presence of substances of this kind in quartz sand is perhaps, at least in some cases, the cause of the occurrence of so-called ' quick-sand.' Occasionally smaller or larger regions of sand on a beach do not show that firmness which is usually observed. They become more or less liquid when one treads upon them, or when they are stirred with a sufficiently high mechanical force. Such quick-sands have not been investigated so thoroughly that one can say whether their behaviour is always caused by the same conditions or not. One has generally assumed that they are due to a kinetic phenomenon, the sand being traversed by streams of water and thus being brought into a mobile state. But there is also another possibility; the sand might contain clay or organic substances, and would, therefore, be transformed into a thixotropic mass that would liquefy when being mechanically agitated. It would be easy to distinguish between these two possibilities. In the first case the sand would only behave as a quick-sand in its natural surroundings : it would not do so if a small sample was investigated in a test tube. But if quick-sand were thixotropic this behaviour would be just as readily observed in a small sample. We found a specimen of quick-sand from Knott End, Fleetwood, which was distinctly thixotropic. It contained 2 per cent of fine clay, whereas normal sand, taken from the same neighbourhood, contained only a small percentage of this clay. The quick-sand lost its thixotropic properties when the clay was elutriated in a suitable way, and regained them when the clay was added again.

Moulding sands seem to be very similar

systems. They are plastic and also thixotropic, owing to a certain percentage of clay or of iron compounds, which have a similar effect to clay.

Another natural occurrence probably caused by thixotropy is correlated with Solnhofen slate, a substance already mentioned. In this slate most beautiful petrifactions are found. It is surprising that animals like jelly fish, which are most delicate and apt to putrify, are found petrified, their patterns being excellently preserved. One is obliged to assume that the period during which the process of modelling took place was very short. Other facts point in the same direction; petrifactions of ammonites are found where the impression of the shell standing upright is seen next to one of the same shell lying flat.

Geologists have come to the conclusion that probably the following conditions prevailed when these petrifactions were formed. The animals were deposited on a shallow beach which was often flooded by the sea. They were covered by a fine dust, blown from neighbouring downs; this dust gave, with the sea water, a sludge in which the forms of the animals were impressed. This sludge must have solidified very rapidly or else the patterns of such perishable animals could not have been preserved. Why this happened is not readily understood. It is most likely not due to quick drying; for in this case one would have expected to find many cracks, and they are only rarely found.

Now, all that is enigmatical could be easily explained if this mixture of dust and sea water had been a thixotropic paste which solidified quickly without losing water. Since thixotropy and plasticity are generally found united in the

same system, such a strongly thixotropic sludge might be expected to be strongly plastic. These assumptions proved to be true. Powdered Solnhofen slate, mixed with sea water, forms masses which are remarkably thixotropic in a wide range of concentrations and which may solidify very rapidly. They are strongly plastic, like putty, and are, therefore, very suitable for· moulding purposes.

This slate contains besides about 95 per cent of CaCO₃ a small percentage of clay ; probably this clay acts in the same way as in quick-sands and moulding sands. Products formed in a similar way as Solnhofen slate often seem to excel as to thixotropy. Bentonite, a mineral found in the U.S.A. and Canada, produced from a disintegrated volcanic dust, gives strongly thixotropic mixtures with water and aqueous solutions. Probably deposits like loess will also behave in this way.

Some thixotropic gels and pastes show the surprising property that, although they are liquefied by strong shaking or stirring, they solidify much quicker under a gentle and regular movement than by doing so spontaneously. If a paste of finely powdered gypsum in water, which, left to itself, solidifies in about ten minutes, is rolled gently to and fro between the palms of the hands, it sets in a few seconds. The same effect may be produced by regular tapping. This may be shown with another example. A $V_2O_5$ sol, containing a small amount of a suitable electrolyte, for instance, a Li-salt, manifests this property in a marked degree.

This phenomenon, which has been called 'rheopexy,' is much less frequent than thixo-

tropy. It was discovered only a short time ago, and one does not yet know all the conditions necessary to produce it, but in any case the size of the particles seems to be important. A freshly prepared $V_2O_5$ sol, with extremely small particles not visible under the ultra-microscope, and containing a suitable amount of electrolyte, is thixotropic and not rheopectic, i.e., it solidifies reversibly spontaneously, but this solidification is not influenced in any way by gentle movement. But if the sol grows old in the course of years, it contains large needles and fibres beautifully visible under the ultra-microscope. Such a sol is strongly rheopectic under conditions which are absolutely ineffective with a fresh sol. The latter may also be aged artificially by heating it in a suitable way for some time to a temperature of about 70° C. The size of the particles is only one of the factors of influence, the nature of the electrolyte added is also important; only a few are active, for instance, Li-salts and acids. Suitable electrolytes all seem to be alike in producing strongly hydrated layers on the surface of the $V_2O_5$ particles. There also appears to be a correlation between rheopexy and the possibility of coagulating a colloidal solution by stirring, just as there is quite generally between making a sol thixotropic by adding a small amount of electrolyte, and coagulating it by adding a larger amount. In any case, the effect of gentle movement seems to consist in allowing larger particles to approach each other sufficiently closely.

By introducing another kind of mechanical force, namely, ultra-sonic waves, one is able to compare thixotropic and rheopectic behaviour

in a very characteristic way. Ultra-sonics may be produced by a piezo-electric quartz plate, lying between electrodes and vibrating in resonance under the influence of an alternating electric field. The whole arrangement has to be isolated well and is, therefore, immersed in an oil bath. In the experiments which are going to be discussed, the plate vibrated with 200,000 cycles per second, i.e., the frequency was more than ten times that of the shortest audible acoustic waves. As soon as the plate commences to vibrate the oil rises above the plate, forming a fountain. In order to let ultra-sonics enter other systems, one only need dip the latter into this oil fountain. That this is actually the case can be shown by stationary longitudinal waves, which may be produced in a liquid and which manifest themselves there in the well-known dust-figures of Kundt : a thick-walled capillary, filled with a suspension of fine quartz powder, is dipped, with its slightly U-shaped middle part, into the oil fountain. It was shown by Bondy and Soellner that in a few seconds beautifully regular striations are formed in the liquid, the distance between two of them agreeing with the half wave length calculated from the frequency and the sound velocity in the liquid. The quartz particles are accumulated in the antinodes, a phenomenon which may readily cause coagulation, partly owing to the increase in concentration, partly owing to the fact that the particles travelling with different velocities are driven together. Thus striations and, therefore, also coagulation may be called forth by ultra-sonics of fairly weak energy.

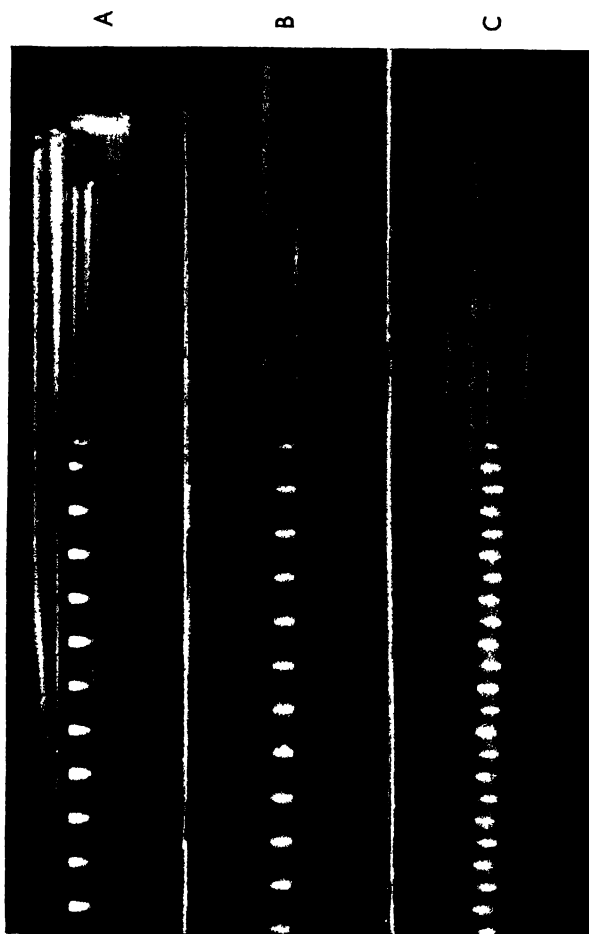I would like to mention, in parenthesis, that

PLATE I.—Stationary Wave-Patterns.

A. Toluene in Water.    B. Quartz in Water.    C. Toluene and Quartz in Water.

these striations obviously represent a special kind of structure in a disperse system ; one may not say ' colloidal system,' because these striations do not appear if true colloidal solutions are used in these experiments.   The size of the particles must exceed a certain minimum of about .5 $\mu$ or else they do not scatter the acoustic energy sufficiently strongly and are not accumulated in the nodes or antinodes.   If the density of the particles is larger than that of water, the latter are accumulated in the antinodes, like the quartz particles (Plate I B) ; but if the particles have a density smaller than water, they are accumulated in the nodes, as is the case with emulsions of benzene, toluene, etc. (Plate I A).   Thus one is able to separate particles of different density : if a mixture of an emulsion of toluene and of a suspension of quartz are treated in this way with ultra-sonics, the droplets of toluene accumulate in the nodes, the particles of quartz in the anti-nodes (Plate I C).

Ultra-sonics with an energy exceeding a certain limit may not only have a coagulating action in places where stationary waves are produced, they may also act as a dispersing agent emulsifying, for instance, one liquid in another.   It has been shown recently that this emulsification is due, in many cases, to so-called cavitations.   The acoustic wave consists of periodical compressions and expansions, and during the expansion phase cavities may be formed in the liquid, provided the energy is high enough.   When these cavities collapse enormous mechanical effects may be produced ; emulsification will be such an effect, when the cavities appear close to the interface of the two liquids.   They are, indeed, preferably formed

there. Since these cavities do not collapse in vacuo the presence of gas is essential for the emulsifying action of the ultra-sonics. Now, this dispersing action of ultra-sonics is likewise conspicuous in their action upon thixotropic gels. On applying ultra-sonics of high energy, a liquefying action is specially obvious. If a sealed-off tube containing a thixotropic $Fe_2O_3$ gel or a paste of gypsum is dipped into the oil fountain with the one end not filled by the colloidal mass, the latter liquefies in a short time and flows down the walls of the tube. This experiment does not work if the colloid has been carefully evacuated. As in emulsification, one is dealing with a liquefication by ultra-sonics, due to the mechanical effects produced by cavities collapsing.

If, on the other hand, weak ultra-sonics act upon a rheopectic suspension of gypsum which has previously been liquefied by intense shaking or by strong ultra-sonics, it is solidified almost immediately. This is brought about by the accumulating and coagulating effects of ultra-sonics ; the particles are driven together in a similar way as is the case when the liquid is gently rolled or tapped. Ultra-sonics are specially effective in this respect, acting so quickly that it is difficult to remove a paste of gypsum, liquefied by strong ultra-sonics, in a liquid form from the oil fountain : the short time, the very last moments, in which the action of the ultra-sonics is fairly weak, is generally sufficient to cause solidification. The collapsing cavities only act momentarily, so to say, whilst coagulation does so more permanently.

True colloidal systems containing ultra-microscopic particles such as rheopectic sols

of $V_2O_5$ cannot be solidified by ultra-sonics; as mentioned already, the latter only accumulate and coagulate particles exceeding a certain size.

The structures which have interested us so far are not of a high degree of symmetry and, therefore, very different from systems of strict regularity, as they are known in crystals. However, one also knows colloidal systems of great regularity similar to those called 'liquid crystals' or 'mesophases,' and they occupy a position intermediate between the isotropic state of liquids or amorphous solids and the anisotropic state of crystals. Since the optical properties are specially characteristic and are mostly used to distinguish these systems, I must discuss them briefly. If a prism is cut out of an isotropic material like glass, and a beam of monochromatic light passes through it, one has the well-known behaviour, first investigated by Newton. But if the prism is cut out of a crystal, for instance, a uniaxial crystal, the optical axis lying normally to the cross section, the beam of light is split into two beams which are polarised normally to each other. The ray whose vibrations are parallel to the optical axis is called the extraordinary ray, the other one having its vibrations normal to the optical axis is called the ordinary ray. This phenomenon is known as double refraction. The degree of double refraction is measured and expressed quantitatively as the difference of the refractive indexes of the two beams :    $Do = n_e - n_o$.

Double refraction is called positive if the refractive index of the extraordinary ray is larger than that of the ordinary ray, negative if the latter is larger.

This experiment is not readily done with a crystalline prism, because one does not often find different sorts of crystals large and clear enough to be cut into a prism, but it is readily done with many colloidal solutions. If a fairly old $V_2O_5$ sol is filled into a glass tube having a triangular cross section, and a beam of monochromatic light is passed through it, the liquid behaves as isotropic as long as it is at rest. But as soon as the liquid is made to flow, it becomes anisotropic: the beam is split up into two, polarised normally to each other. This is stream double refraction, a phenomenon shown also by many other colloidal solutions, for instance, of dyestuffs, like benzopurpurine, cotton yellow, etc. Before entering upon the cause of this phenomenon, another experiment may be mentioned which is especially fit for demonstrating stream double refraction. If a glass plate is brought between two crossed nicols, the field of vision remains dark. If, instead of a glass plate, a plate of a uniaxial crystal, cut parallel to the optical axis, is interpolated between the two nicols, at an angle to the planes of polarisation, the field of vision lights up, because a component of the polarised light is able to pass through the second nicol. This experiment can be easily performed with a colloidal solution which is stream double refracting. A tube of rectangular cross section, filled, for instance, with an old $V_2O_5$ sol, is brought between two crossed nicols. As long as the liquid is at rest the field of vision remains dark, as soon as it flows at an angle to the planes of polarisation it lights up. In a less accurate form it is done thus : if a cell containing such a sol is brought between the nicols and the

liquid is stirred, the field of vision lights up because there are always some currents fulfilling the conditions necessary to allow a component of the light to pass through.

Other external forces may produce the same effect, for instance, a magnetic field or an electric current. If a cell, again containing an old sol of an iron hydroxide FeO.OH, is brought between the poles of a sufficiently strong magnet, the lines of force being normal to the direction of the light passing the cell, the field of vision lights up as soon as the magnetic field is set up.

In all these cases we are dealing with an orientation of particles as may be seen directly in the ultra-microscope. Taking stream double refraction as an example, as long as the colloidal solution is at rest the particles are distributed at random in lively Brownian movement. Even if they are, for instance, rod-shaped, consisting of small crystals whose optical axis is parallel to their long axis, no double refraction need be observed : the effect of particles lying in a direction favourable to a component of the light passing through, is compensated by particles lying unfavourably. The whole system behaves as isotropic. But, if the liquid flows, rod-shaped particles are orientated practically parallel to the lines of flow. A sheet of liquid whose particles are lying to a high percentage with their long axis and, therefore, with their optical axis parallel to the lines of flow, would behave like a plate of a uniaxial crystal cut parallel to its optical axis. Such an orientation only occurs if the particles are anisometric, i.e., rod- or plate-like ; spherical particles would never be orientated in this way. Thus stream double refraction

caused by the orientation of non-spherical particles is always the surest method for determining the shape of colloidal particles. Even if the particles have no intrinsic double refraction, but are amorphous and optically isotropic, a certain double refraction, the so-called rod- or plate-double refraction, is produced, provided the particles are non-spherical and are orientated in a regular way.

In magnetic double refraction, conditions are similar, but not equally simple owing to the susceptibility being dependent on the crystal structure. Even spherical particles may be orientated by a magnetic field, from a theoretical point of view, although this is hardly probable under normal conditions.

In concentrated colloidal solutions containing, for instance, rod-shaped particles, orientation of this kind may take place spontaneously. If some concentrated old $V_2O_5$ sol is brought between the nicols of a polarisation microscope, care being taken to avoid evaporation, flame-like double refracting spots, so-called ' tactoids,' are formed in course of time (Plate II A). They are not crystals, for they may be dispersed by stirring the liquid, moving the cover glass on the objective, and they form again in a certain time. These spots may also be seen under the ultra-microscope if the specimen is illuminated with light coming only from one side, striking the long axis of the particles normally (Plate II B). Under the ultra-microscope one may see that the particles in these groups are not at rest, but have a slow rotatory Brownian movement, a further evidence that we are not dealing with crystals.
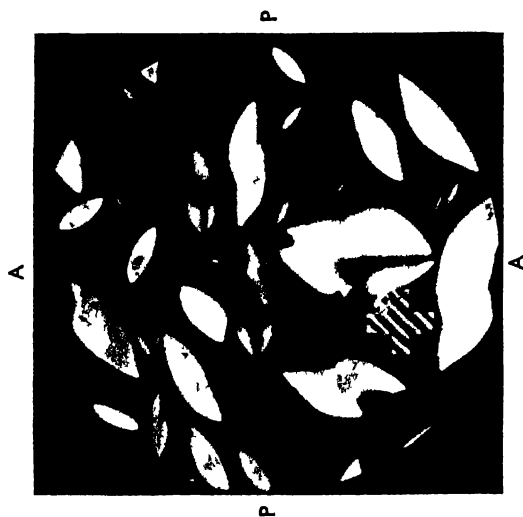
PLATE II

A  Tactoids under the polarisation microscope
   P P  position of polarising nicol
   A A  position of analysing nicol
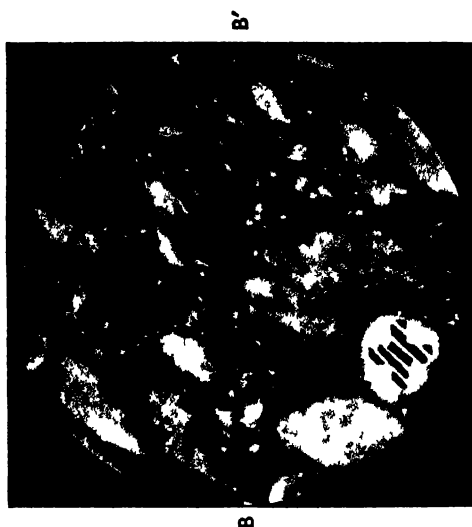
B  The same tactoids under the ultramicroscope,
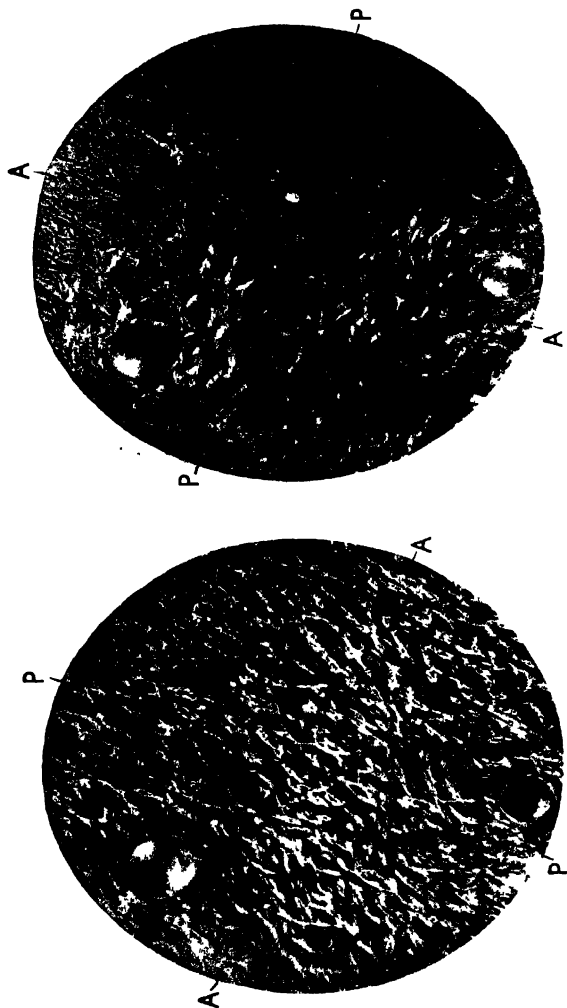   illuminated with light coming from one side
   BB    Direction of light

PLATE III —Tactoids of $V_2O_5$ and benzopurpurine under the polarisation microscope, using a sensitive gypsum plate. Streaky tactoids : $V_2O_5$, round tactoids : benzopurpurine. PP. Position of polarising nicol. AA. Position of analysing nicol. The dark regions are yellow, the light ones reddish brown to green.

Here again the question arises most forcibly, what is the nature of the far-reaching forces of attraction acting here ? One important property of these forces can be found by investigating these tactoids : the forces are distinctly specific. This may be proved by the behaviour of sols containing particles of different kinds, both forming tactoids. Both sols of $V_2O_5$, if sufficiently old, and of the dyestuff benzopurpurine, are able to produce tactoids. In both cases the particles carry a negative electrical charge ; therefore, the danger of their being mutually coagulated is not so very great, and it is possible to adjust the concentration of particles and electrolytes so that in the mixed sol tactoids of $V_2O_5$ and of benzopurpurine are formed, both containing only particles of one sort. They may be distinguished because the $V_2O_5$ tactoids are positively double refracting, those of benzopurpurine negatively double re-fracting. (Plate III ; for instance, the axis of the smaller benzopurpurine tactoid runs from the right hand side above to the left hand side below ; it is mainly yellow, whereas tactoids of $V_2O_5$ lying in the same direction are reddish brown to green.) This specificity shows that we are probably dealing with molecular forces, i.e., forces of van der Waals'.

I have not been dealing with the technical application of these phenomena, nor with their biological significance. But I might mention that protoplasm frequently shows a thixotropic behaviour, and still more frequently is in a state which may readily be made thixotropic by small changes in the protoplasmatic com-position. Structures similar to tactoids have also been observed in the processes going on in

living cells. Solutions of myosine, a protein contained in muscle fibres, are both thixotropic and stream double refracting, and the double refraction in muscular fibres is due to the presence of this substance and most probably to an orientation as discussed above.

[H. F.]

MAJOR CHARLES E. S. PHILLIPS, O.B.E., F.R.S.E.
*Secretary and Vice-President, in the Chair*

SIR WILLIAM BRAGG,
O.M., K.B.E., M.A, D.Sc., Pres.R.S., M.R.I.
*Fullerian Professor of Chemistry*

# THE ELECTRIC PROPERTIES OF CRYSTALS (II)

THE recently acquired and more intimate knowledge of the manner in which the molecule is built into the solid* allows a further insight into the problems of the dielectric constant and of the electric properties of the crystal. These problems have been widely studied of late because of their extreme importance in industry, especially in the transmission of electricity at high voltages. They are of remarkable interest also for their applications in radio telephony, where great use is being made of the piezo-electric property in the production of oscillators of known high frequency, and in the construction of various types of microphone.

A model of the resorcinol structure was used during this discourse to illustrate the polarity of the crystal due to the peculiar arrangement of the dipoles contained in the molecule, and the variation of the electrical polarity with heat, that is to say the pyro-electric effect.

The structure of the Rochelle Salt crystal was

* Electric Properties of Crystals (I) Roy. Inst Procs. Vol. xxix, p. 225 1936.
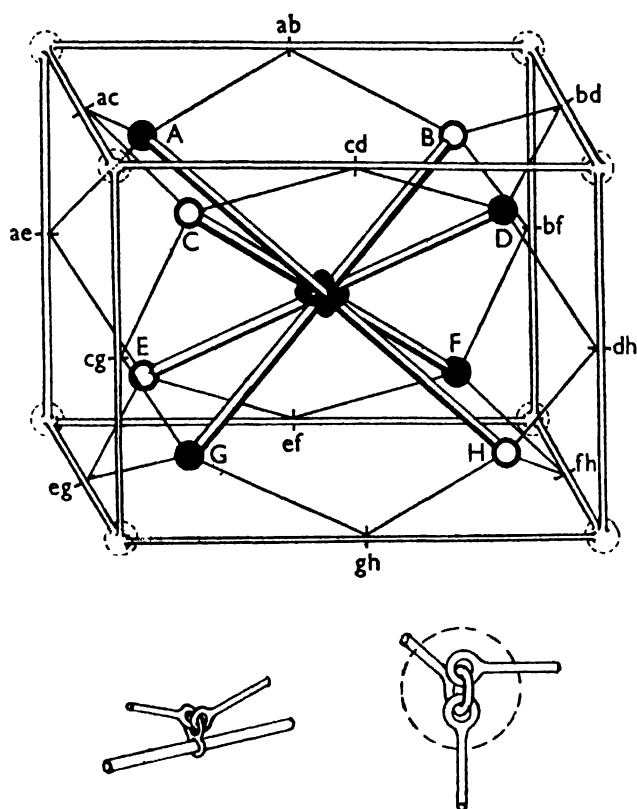
FIG. 1.—This figure shows the form of the model of Rochelle Salt referred to in the text. The linkages are illustrated separately and enlarged. That on the left is the linkage at the middle point of a side such as the point **ae** above. That on the right is the linkage of each corner.

also illustrated by a model, but only in part as its complexity had not yet fully yielded to analysis. Enough was known, however, to show how the arrangement of the molecules in the crystal unit connected the exertion of strains in certain directions with the appearance of electrical

polarities at right angles to those strains. The model is shown in Fig. 1.

In this model of the unit of structure of Rochelle Salt, each of the four molecules is represented by a rod with a black sphere at one end and a white one at the other, the difference of the ends being an indication of polarity: the four are AH, BG, CF and DE. The arrangement of the four, which are shown crossing at their centres, gives the unit a rhombic-bisphenoidal character, this being the crystallographic description of a crystal which has no planes of symmetry, but has three axes of digonal symmetry at right angles to one another. This class of crystal symmetry is illustrated in Figs. 2 and 3. The three axes (not drawn) pass each through the centres of two opposite faces of the rectangular framework, which represents the form of the unit cell. Each end of a molecule is connected to the middle points of three adjacent edges of the cell. These connections represent atomic and molecular linkages which have not yet been investigated by the X-ray methods. Thus the end A of the molecule is connected to three points denoted in the figure by ab, ac and ae : B is connected to ab, bd and bf and so on. In the model the framework of the cell, and all the junctions such as those just mentioned are made of stiff wires with the exception of a selection which must be made of elastic bands so that the cell is capable of distortion as a whole. For instance, such a selection might be the set of bands joining the ends of the molecules to the middle points of the vertical edges of the cell, ae, cg, dh, bf. The four rods that cross in the centre are tied together there by an elastic band; they have a certain amount of free play about the

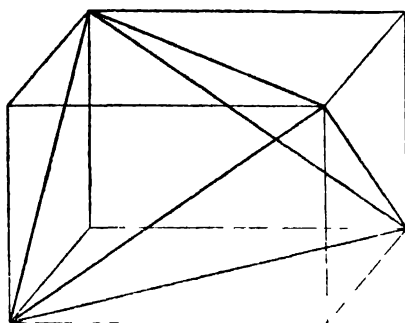FIG. 2.—The tetrahedron inscribed in the rectangular frame shows the essential features of the rhombic bisphenoid. If the frame is turned through 180° about any axis parallel to an edge and passing through the centre the new position of the tetrahedron coincides with the old.
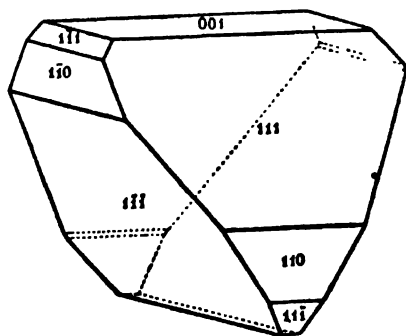


FIG. 3.—The crystal of tartar emetic is an example of the rhombic bisphenoid. The illustration is taken from Fulton's Crystallography Vol. I, p. 230.

centre. The eight corners of the cell are loosely jointed.

If now the cell is distorted by a shear whose plane contains the vertical in the diagram, all the positive ends of the molecules will ascend or descend together, while the negative ends do the reverse. This may be seen by considering a set of points such as C and G joined to cg, and B and F joined to bf. The set resembles lazy tongs. If the vertical edges containing cg and bf are pushed together (the edges containing ae and dh necessarily move apart at the same time) the molecule ends C and B rise while G and F fall. At the same time, A and D fall while E and H rise. All the positive ends go one way and all the negative ends go the other. The crystal is polarised in the vertical direction. It is this piezo-electric action which is utilised in practice.

The piezo-electric properties of quartz are also much used. It is not easy, however, to show by model or diagram, the connection between the possession of this property and the actual structure of the crystal, although the latter is well known in the case of quartz and imperfectly in the case of Rochelle Salt.

The very interesting properties of Rochelle Salt are under comprehensive investigation by Dr. Scherrer and his colleagues at Zurich. A good account is given by René David.*

Various experiments were shown during the discourse in order to illustrate the dielectric and other electric properties of crystals, and the applications of these properties. Through Dr. Rayner, the National Physical Laboratory contributed a number of illustrations of the vibratory properties of

* Helvetica Physica Acta Vol. viii, p 431. 1935.

quartz, obtained during the late Dr. Dye's researches on the use of such vibrations in radio transmission. Also Dr. Kaye, of the same Laboratory, lent an instrument showing how the piezo-electric properties of quartz were used in determining the length of resonating columns of air or other gases at different temperatures, and the Building Research Station showed how they could be made to register the internal pressures of piles when being driven. Messrs. Rothermel supplied a number of demonstrations of the uses of Rochelle Salt in microphones and loud-speakers.

[W. H. B.]

## Friday, May 22, 1936

SIR RICHARD PAGET, Bart., F.Inst.P.
*Manager and Vice-President, in the Chair*

E. N. DA C. ANDRADE, D.Sc., Ph.D., F.R.S., M.R.I.
*Quain Professor of Physics in the University of London*

## WHIRLPOOLS AND VORTICES

IF asked what is meant by a vortex, the average cultured man probably turns his thoughts to the hollow funnel that is often formed by the water running out of a bath or basin. He has observed that the water surrounding this funnel is going round and round, the speed of rotation being greater where the funnel is smaller, and that little bits of floating matter are rapidly, as the saying is, drawn into the vortex. He probably also thinks of Edgar Allan Poe's striking story *The Descent into the Maelstrom*, with its vivid description of the boat sucking into the shining shaft. Unfortunately, just as the waves of the sea, which the average man thinks of as typical waves, are exceedingly bad ·examples of what the physicist generally means by a wave, so the whirlpool, or bath vortex, is an exceedingly bad illustration of what the physicist technically calls a vortex. Nevertheless, the discussion as to why it is a bad illustration may give us a good starting point for a talk on vortices.

The essence of the physicist's and mathematician's vortex is rotation, but rotation of a definite kind, which may be absent even when a liquid is going round and round with its individual particles tracing out closed curves. Let us con-

sider the case of a soldier walking round and round an enclosed space. If he were told to be on the look out for an alarm from a particular direction, say the north, and always to face in that direction, he would complete each circuit without ever having turned on himself. If, however, he were equally suspicious of all directions, he would always face outwards, and at the end of a circuit he would have rotated once on himself, just as the moon does on her journey round the earth. We always see the same face of the moon, because she turns once on her axis in the course of her circuit in her orbit. It is the turning of fluid particles about an axis through the particle that constitutes what is technically known as vortex motion to the physicist. Within a vortex any particle is in the course of turning on its own axis.

Lines giving us the paths of the centres of fluid particles cannot of themselves tell us whether the motion is vortex motion or not. The simplest case is where the particles all move in circles. Here we may have either vortex motion or vortex-free motion, as we may now proceed to consider.

Let us first of all take a liquid in a vessel which is set rotating on its axis at a constant rate. At first the liquid moves only where it is in contact with the walls, but very soon, owing to viscosity, the liquid rotates as a solid mass. The centrifugal force, which becomes greater as we proceed outward from the centre, tends to throw the liquid outwards and keep it up at the edge; gravity tends to keep down the liquid to a constant level; as a result of the joint action of these two forces the liquid surface takes up the form of a paraboloid of revolution. It is well known

that a parabolic mirror brings a parallel axial beam of light to an exact focus. A dish of mercury rotating on its axis can therefore be used as an astronomical telescope, and some twenty-five years ago Professor R. W. Wood constructed a telescope of 20 inches diameter on this principle. By varying the speed the focal length can be altered, and he used focal lengths of from three to twenty feet. I have here a shallow vessel containing mercury which can be set into rotation, and act as a mirror, of adjustable focal length, which throws a clear image of a bright object. [EXPERIMENT.]

It is clear in this case that every particle of liquid rotates on its axis, once in a revolution, so that here we have true vortex motion. Suppose now that instead of rigid walls we let our cylindrical vortex have liquid walls, which we can do approximately by rotating a suitably shaped paddle, as shown, in a large cylindrical vessel. We then have a central vortex, and the liquid surrounding it also moves in circles, but in this surrounding liquid the speed of rotation diminishes as we go outwards, until at the walls it is at rest.

In these outer regions we find that the particles of liquid are moving, approximately speaking, without rotation, always facing the same way. We have in these regions, then, a whirlpool which is not a vortex. The surface here has a different shape. The great experimenter and physicist, W. J. M. Rankine, who first considered an idealisation of this case, called the core a forced vortex, and the surroundings a free vortex, but this terminology has disappeared.

Plate I shows such a case : in the middle we have a cup-shaped surface, corresponding to a

PLATE I

true vortex, like our mercury mirror; outside, the surface is shaped like the outer part of a bell, corresponding to the bath whirlpool.

The mathematicians have introduced a term which they call the circulation. If we go round any small closed path, divide up the path into small bits, multiply the length of each bit by the velocity in the direction of the path, and add all these products (or, as they say, form the integral round a closed path) the result is called the circulation. If we do this anywhere within a circular vortex, and divide by the area of the circuit, we get a fixed value, which we call the strength of the vortex. Any little circuit outside the vortex, *i.e.*, not including the vortex, gives a zero result—the circulation is zero. If, however, the circuit includes the vortex, we get the same value for the circulation, whatever the size of the path.

Now the electric current furnishes an exact analogy with this state of affairs. Within the wire conveying the current the lines of magnetic force are circles, the force increases with the distance from the centre, and the line integral of the magnetic force round any unit area is the same everywhere. Outside, the lines of magnetic force are still circles, but the force decreases inversely as the distance from the centre. The line integral of the magnetic force round a circuit is zero if the circuit does not embrace the current, and, if it does embrace it, has a value which is the same for all circuits. The magnetic force due to a straight conductor corresponds, then, exactly to the velocity due to a cylindrical vortex. We shall come back to this analogy later.

Let us now return to our water whirlpool. A particularly simple case occurs if the motion of

the liquid is irrotational in circles. We can then show that the velocity, in a perfect fluid, is inversely as the distance from the centre. Owing to the centrifugal force it will tend to clear away from the axis, and we have formed what is known as a hollow vortex, which reminds us of the bathroom vortex. Strictly speaking, it is not a vortex at all. The cylindrical vortex has, as we have seen, a basin-shaped surface : the irrotational circulation has a funnel-shaped surface. I prefer to call it a whirlpool. It is difficult to realise experimentally since the rotation has to be greatest at the centre. The nearest thing to it is probably given by rotating a solid cylinder in the middle of a very large vessel.

It is evidently some whirlpool of this sort that Poe had in mind in his story. I cannot find that he had any justification for the scientific " fact " which is the peg on which the whole story hangs, namely that " a cylinder, swimming in a vortex, offers more resistance to its suction, and is drawn in with greater difficulty than an equally bulky body, of any form whatever." Looking at the story again for the first time since I read it as a boy, I see that he quotes as his authority Archimedes *De Incidentibus in Fluido*, book 2. Quite apart from the bad Latin there is nothing at all about moving liquids anywhere in this work, which deals entirely with floating bodies at rest in a liquid.*

* The title of Archimedes' work is περὶ τῶν ὕδατι ἐφισταμένων usually rendered in Latin versions as *De Insidentibus Aquae*. The second book deals fully with the conditions of stability of a right segment of a paraboloid floating at rest in a fluid. It is a masterly piece of work, but there could hardly be anything less relevant to Poe's story. It must clearly have been quoted at random, for effect, probably as the only book on fluids of which Poe had heard.

A great authority, having described the " hollow vortex," has stated that "such funnels are often to be seen during the emptying of a bath." There is, however, an essential difference. In the bath whirlpool conditions are not steady : water is pouring all the time from the outer parts into the walls of the hollow space. Here we have a new factor coming in—the downward component of momentum of the water. This makes the bath whirlpool a very complicated phenomenon, and, as a matter of fact, it has never been
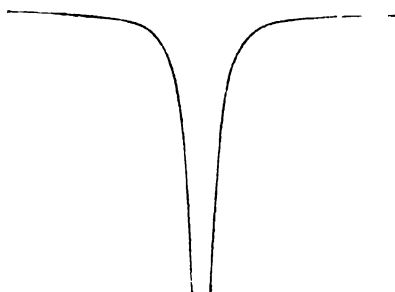


FIG. 1.

properly worked out. In general shape, however, it resembles closely the "hollow vortex" of the text books. (Fig. 1.)

When we come to consider how the bath whirlpool originates, we have to take into account what is known as the conservation of angular momentum. If, namely, a particle which is circulating round an axis alters its distance from that axis, it must change its speed so that the product of distance by speed remains the same—the closer it comes the faster it must move, and, still more, the greater must be the number of times round

per second, since, the distance being less, the angular velocity would have to increase if the velocity had to remain the same. There are many illustrations of this in ordinary life. A dancer who wants to spin very fast, for instance, gets up as great a spin as possible with her arms out ; when she drops them the rate of spin increases. This can be simply shown by the aid of a small stool which is free to rotate. If a man stands on such a stool with a weight in each hand and his arms held out horizontally, and is given a very slow spin, then, on dropping his arms, which brings the weights nearer the axis of rotation, his rate of spin becomes very large. [EXPERIMENT.] In the case of the whirlpool this means that any particle which is moving round slowly near the edge of the vessel must circulate much more quickly as it approaches the axis of the exit pipe, so that in this way it acquires the speed necessary to keep the funnel open. The greater the movement near the edge of the vessel, the sooner the whirlpool will form.

This brings us to a question that is often asked. It is generally believed that in a given bath or basin the whirlpool goes the same way, clockwise or counterclockwise, every time the vessel is emptied, and it is often asked if this is due to the rotation of the earth. Before trying to explain the facts, we must ascertain what they are. I find that it is possible to suppress the whirlpool completely if the water is allowed to become quite still before the plug is withdrawn: a cylindrical vessel with a central hole, where the conditions are more symmetrical than in the ordinary basin, can be emptied without any vortex if sufficient care is taken to avoid any preliminary movement of the
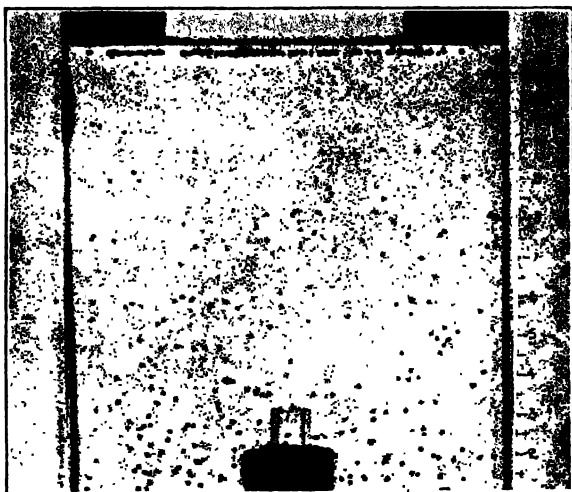
**A.** Water running out of central aperture (at bottom of picture) without preliminary rotation.
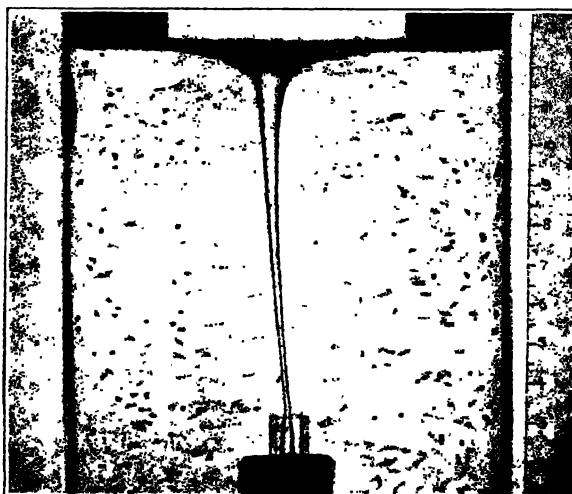


**B.** The same, with a slight preliminary rotation. The dots are suspended particles, of density equal to that of water. In (B) the rapid rotation near the axis makes this appear blurred.

PLATE II.

water. (Plate II, A and B.) If, on the other hand, any slight rotation is given to the water before the plug is withdrawn, a splendid whirlpool forms almost immediately, the sign of the rotation, clockwise or counterclockwise, being that of the preliminary movement. The increase of the rotation from the slight movement at the edge to the vigorous spin in the funnel is due, as just explained, to the conservation of angular momentum. I attribute the whirlpools ordinarily formed, then, to the fact that there is always, in ordinary circumstances, a slight rotation in the water before the plug is withdrawn. If the rotation is always in one direction with a given basin it must be because some chance feature, such as the way the hands are withdrawn or the influence of the shape of the basin, is favourable to the imparting of a circulation in one direction.

Does the rotation of the earth, then have no influence on the movement of the liquid towards a central orifice ? Actually, of course, everywhere except on the equator it must have a component about a vertical axis, and the effect of this component is that the movement, instead of being purely radial, will be in spiral lines, as shown, in a very exaggerated manner, in Fig. 2. The effect, however, can be calculated to be so small that it requires the greatest care to establish it experimentally with a vessel the size of an ordinary basin. An Austrian scientist named Tumlirtz did, however, by taking sufficient trouble, succeed in showing the effect. He allowed water between two circular glass plates, about 5 feet in diameter and 2 inches apart, to flow into a central hole. The water was fed from the edges of the plates, where it was coloured in places so as to make the

lines of flow visible, and the greatest precautions were taken to avoid any initial movement of the water and any temperature effects.   The flow was so slow that at the edges of the discs the water was moving in at a rate of about 1 inch in 25 minutes.
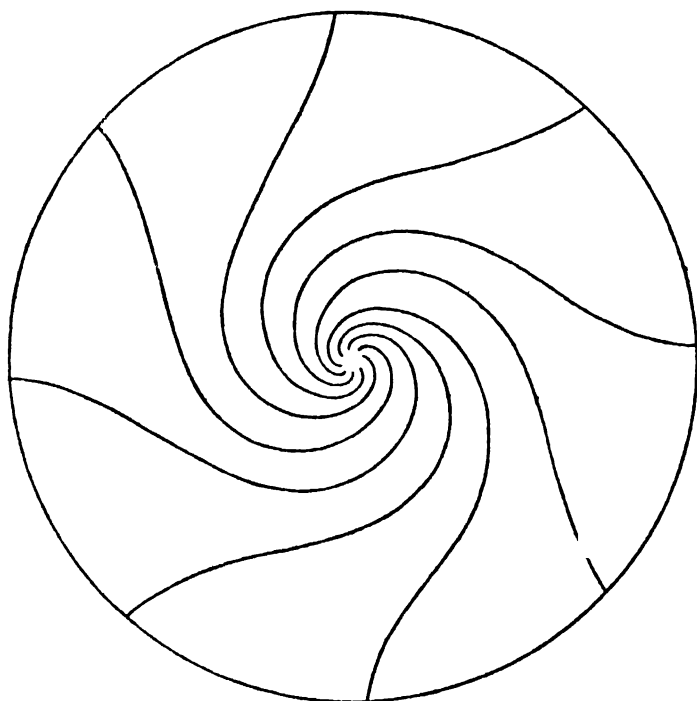


FIG 2

With this arrangement he obtained spiral lines of flow that twisted in the direction to be anticipated from the rotation of the earth, although the work was not sufficiently accurate to allow the speed of rotation of the earth to be deduced, which should be theoretically possible.

A    Clockwise vortex at turbine intake  Arapuni, New Zealand
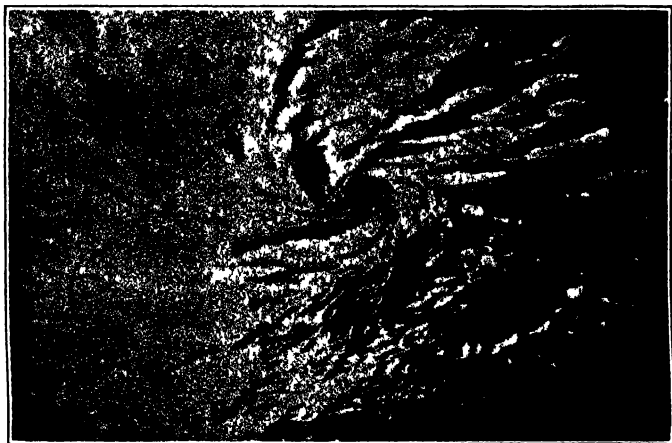


B    Anti-clockwise vortex at Loch Treig, Scotland
*(By courtesy of Messrs  Glenfield and Kennedy, Ltd )*

PLATE  III

The conditions for making evident the rotation of the earth by central flow are actually realised on a large scale at the intake of the water for water turbines in certain cases, where a large, still sheet of water flows away through a vertical pipe. A whirlpool is often formed in this case, and in three cases in the Northern Hemisphere cited by Mr. Bruce Ball the rotation of this vortex is in the direction to be anticipated from the earth's rotation. In the one case recorded in the Southern Hemisphere the rotation, as seen by a local observer, is in the opposite direction, which is as it should be if the earth's rotation is the cause. The rate of central flow is rapid, which is against the establishment of a marked whirlpool, but on the other hand, the water flows in over a very extensive sheet, which is favourable for the formation of well-defined spiral lines of flow. I have made calculations which show that if the water is fed in from a disc 50 feet or so in radius, then with the kind of velocity that occurs, say 1 foot per second near the edge of an outflow pipe 8 inches across, the spiral lines of flow should have just about the curvature that is actually observed, as illustrated in Plate III, A and B. It seems tolerably certain, then, that in these cases we have a true effect due to the rotation of the earth.

The bath whirlpool has its close analogy in the waterspout, as we might guess from its appearance, which closely resembles that of an inverted waste-pipe whirl. The waterspout probably originates in a low pressure region above the sea, into which the air flows, any slight initial movement changing into rapid rotation by the conservation of angular momentum. Such waterspouts are illustrated in Plate IV. The air flows

away upwards into the low pressure region, just as the water flows downwards, and forms an air whirlpool. The waterspout itself consists not of water, but of a cloud of water drops and droplets. The well-known meteorologist Buchan says in his *Handy Book of Meteorology* (1868): "It is a popular fallacy that the water of the sea is sucked up by them" (waterspouts), "it being only the spray from the broken waves that is carried up into the vortex. This is conclusively proved by the fact that the water poured down on the decks of vessels from waterspouts is either wholly fresh or slightly brackish." The fact is, no doubt, as cited by Buchan, but it does not prove what he says, for if the water were spray it would, of course, be as salt as the sea. Actually the water drops of which the spout is composed are condensed from the saturated vapour above the sea, by the expansion of the air as it flows into the low pressure region, the effect resembling that produced in the celebrated Wilson cloud chamber. This does account for the freshness of the water.

Let us now turn to the true vortex, whose characteristic, as we have seen, is that within the vortex every element has a rotation about its axis, which in the simplest case means that the velocity of circulation increases as the distance from some axis, while outside the vortex there is likewise a circulation, but without local rotation, which in the simplest case means that the velocity of circulation decreases as the distance increases. In the mathematician's perfect fluid, loved for its comparatively simple relations, it would be impossible to start a vortex*, for there is no fluid friction, no

* This is the statement usually made, but my friend Professor L. N. G Filon has pointed out to me that there may conjecturally be cases where it is not true
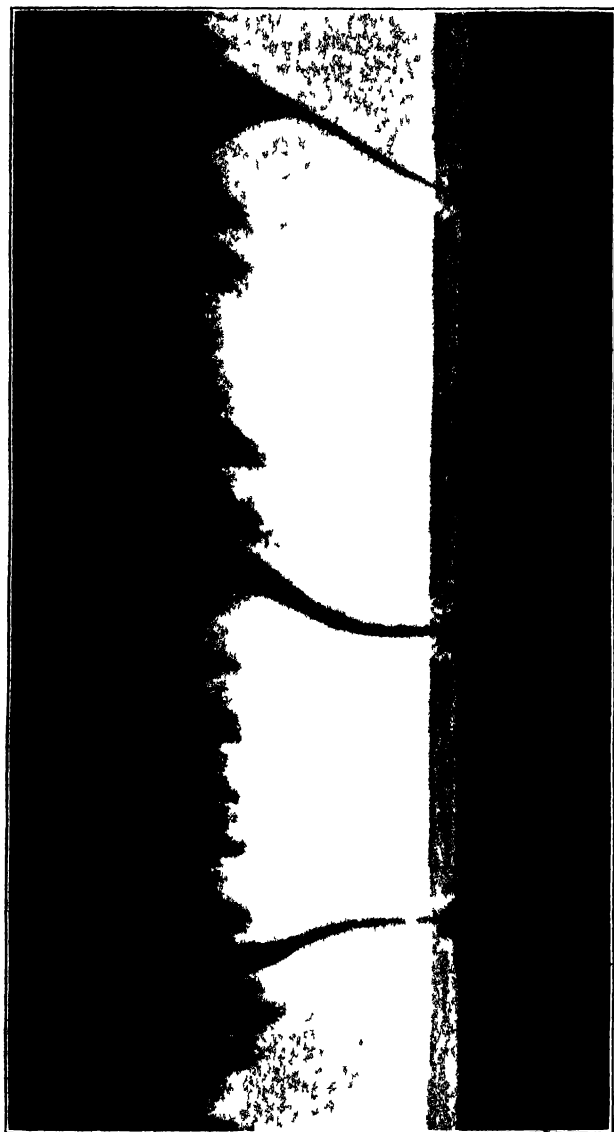
PLATE IV    Waterspouts

(By courtesy of Sir George Simpson, F R S )

viscosity, to allow us to apply a tangential force. For instance, in the flow of a perfect fluid past a cylinder the stream lines curve, but any element of the fluid, while changing its shape and position, does not rotate. On the other hand, if in such a perfect fluid we suppose that a vortex existed, by some act of creation, it would go on for ever : it could travel about, but could never be dissipated, as actual vortices are. It was for this reason that Kelvin made up his theory that atoms were vortices in the ether, which was supposed to behave like a perfect fluid. In those days atoms were supposed to be unchangeable and indestructible, but to-day the fluid ether and the unchangeable atom are equally discredited, and with them the vortex atom has disappeared.

In a real fluid vortices can easily be generated, and, once generated, their motion is gradually frittered away into heat by the action of viscous forces. Vortices usually owe their origin to surfaces on the two sides of which there are large differences of relative velocity—surfaces of discontinuity, as they are called. Such surfaces can be found when two streams flow together in the same direction, from opposite sides of a long obstacle ; or where the fluid streaming past an obstacle is separated from the " dead water " ; or where a fluid streams round or past a sharp corner. A surface of discontinuity can be regarded mathematically as a sheet made up of a number of very small parallel cylindrical vortices, as illustrated in Fig. 3. Such a sheet is very unstable, and easily rolls up, disengaging a large vortex or vortices. A good example is offered by the familiar vortex ring or smoke ring. A rapid jet of air is sent past the edge of a circular hole,

forming a surface of discontinuity with the sur-
rounding air where it issues. This rapidly rolls
up into a ring, as illustrated in Fig. 4, which travels
out, the air surrounding the actual ring being
carried round, just as the fluid surrounding a
cylindrical vortex is. With these rings many
pretty experiments can be carried out, especially
in illustration of their action on one another.
If the smoke be projected through an elliptical
opening an elliptical ring is produced, which
vibrates so as to have its long axis alternately
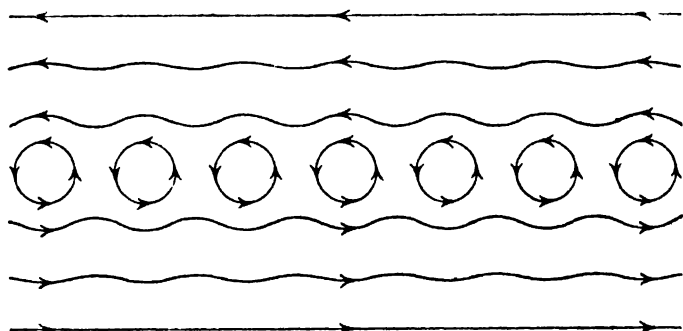


FIG 3.

vertical and horizontal. If one ring be sent after
another it contracts and passes through, the first
expanding to let it pass. If two rings be sent
towards each other from opposite directions they
expand, owing to their mutual action. This is
best demonstrated by sending a single ring against
a plane wall, which can be shown to amount to the
same thing as if there were a second ring, advancing
as would an image of the first one reflected in the
wall as a mirror. Very pretty rings can be pro-
duced by liquid jets, as, for instance, of water
coloured with eosine into water : they break up

in a symmetrical manner with the production of minor vortices. A single liquid drop, falling into a liquid, also produces a vortex ring, as can be demonstrated with a solution of mercuric chloride falling into potassium iodide, the ring being marked by the precipitated mercuric iodide. [EXPERIMENT.]

That the formation of the smoke ring is not due to the friction at the solid boundary of the hole,
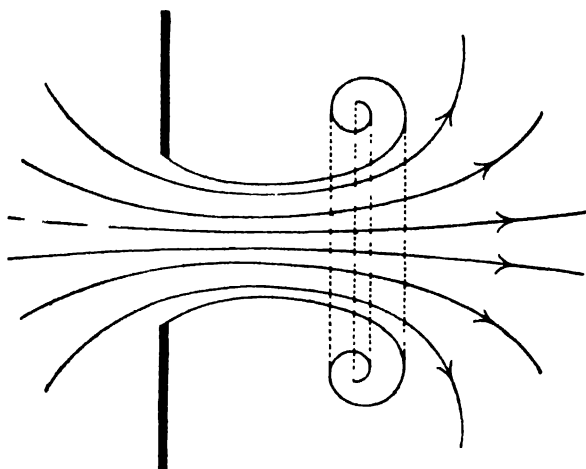


FIG. 4.

as is sometimes stated, can be shown by covering the hole with a sheet of zinc perforated by holes, which means that such friction is taking place right across the jet. The vortex rings produced with the perforated zinc in position are indistinguishable from those produced in its absence, showing that it is the difference of velocity between the boundary of the jet and the still air that causes the vortex.

Another case, of great interest, is the generation

of vortices behind an obstacle, either when it moves through a liquid, or when it is stationary and a liquid flows past, which amounts to the same thing.   This is of practical importance, for it greatly influences the resistance offered by the fluid to the motion of the obstacle : the so-called streamlining of moving bodies is carried out with the object of avoiding vortex formation, as explained by Professor Melvill Jones in his discourse here last year.   If we have streamline motion of a sphere or cylinder through an ideal liquid we have the case contemplated by Coleridge in *The Ancient Mariner* :

> "But why drives on that ship so fast,
> Without or wave or wind ? "
> "The air is cut away before
> And closes from behind."

a condition which leads to motion without resistance.   In the ordinary way, however, the fluid does not close in behind, but breaks up into vortices.   Let us consider the simplest case, the motion of a plate.   In the ideal liquid the flow diverges to each side of the plate, and leaves a region of " dead water " which extends right back, and in which the pressure is the same as it is in the undisturbed liquid.   We have a surface of discontinuity between the dead water and the moving liquid.   With any real liquid the surface is very unstable, and tends to break up into vortices.   A particularly beautiful series of vortices, disengaged alternately to the right and to the left is formed, which has been investigated by Karman, and is called a Karman vortex street.   It is one of the rare cases of a stable vortex system that can be treated by mathematicians.   Needless to say, the

formation of vortices leads to a large dissipation of energy, and is consequently attended by a considerable resistance to motion.

We have already noticed that vortices cannot be set up in a fluid absolutely free from viscosity. For instance, suppose a cylinder rotates on its axis in the fluid, no motion will be communicated, since the fluid will slip over the surface without any friction. Real fluids, even the least viscous, do not, however, slip at the surface of solids. Supposing, then, that some slight viscosity exists, so that slip does not take place, then we find that the lower the viscosity the easier vortex motion takes place. The viscosity exercises a stabilising influence, and enables laminar flow to take place in cases where with a vanishingly small viscosity it would not. An excellent example of this is provided by the case of a liquid between concentric cylinders, fully worked out by G. I. Taylor. If the inner cylinder rotates, and the outer one is fixed, then for very slow speeds of rotation the viscous liquid, say water, moves in concentric layers, like a series of closely fitting tubes turning on one another. As the speed is increased a point is reached when the laminar motion becomes unstable, and breaks up into a series of ring vortices, fitting on the inner cylinder like a series of rubber umbrella rings, and rotating alternately in opposite directions. The lower the viscosity, the lower the critical speed : the viscosity stabilises the laminar motion. [EXPERIMENT.]

Let us now turn to a particular problem with which I have been concerned, and which is traditional to the Royal Institution, I mean that of the fluid-into-fluid jet. This may be exemplified by either a jet of air into air, or a jet of water into

water. One of the most familiar examples, however, is that of coal-gas into air, and ignited· coal-gas at that. This great difference in temperature and composition between the jet and the surroundings complicates matters very much, but makes certain phenomena easy to observe. The most striking of these phenomena is the so-called sensitive flame. A long flame is turned up until it is just on the point of becoming unstable. It will then respond to sounds, especially to those of high pitch, by breaking into turbulent motion. Such flames were shown by Tyndall in this theatre as long ago as January 18, 1867. Recently a good deal of work has been done on them in my laboratory, both by Dr. G. B. Brown and by myself, and we have a general idea as to what happens in the process. [EXPERIMENT.]

To avoid the complications of the flame it is better to work with cold fluids, and to make the motion visible by placing something in the jet. Dr. G. B. Brown has carried out his work with a jet of air into air, marking the jet by introducing a little cold tobacco smoke into it. I have worked with liquid into liquid jets, using sometimes coloured water into water, and at other times heavy organic liquids containing suspended particles. Dr. Brown has followed the periodic motion of the jet by illuminating it stroboscopically, and either watching it or photographing it with a cinema camera, and has found that when to the eye it appears to be turbulent, it is actually moving periodically in time with the disturbing sound. It discharges vortices regularly to right and left, which travel up, accompanying the sinuosities of the main stream. I am able to show you this evening a short film illustrating his observations.

The general phenomenon I will illustrate with a water-into-water jet, which breaks up to sound in a very marked manner. The motion is really periodic, just as with the air-into-water jet. Very remarkable is the way in which the jet responds strongly to certain particular frequencies, as does also the air-into-air jet. This phenomenon puzzled me for a long time, but ultimately I have succeeded in showing that the apparent selective response is not a property of the liquid jet, but of the surrounding tubes and apparatus. These have their own mechanical vibrations, with determined frequencies, and come into resonance if one of these frequencies is sounded. That it is resonances of the apparatus that are made manifest by the selective behaviour of the jet was confirmed by loading the jet-tube and other parts of the apparatus, which shifted the frequencies at which marked response took place.

In investigating the liquid-into-liquid jet I have made use of the hydromagnetic analogy, if I may so call it—the fact, to which I have called attention, that the lines of flow round a cylindrical vortex are exactly analogous to the magnetic lines round a straight axial current, the strength of the magnetic force corresponding to the magnitude of the velocity. The work has not yet been published, but I may state that the whole phenomenon of the sensitive flame is generally explicable along the lines of considering it as a double vortex sheet.

At one time, about three hundred years ago, vortices were considered to be the most important thing in science, for they were supposed by Descartes to carry round the planets, and magnetism and all other physical phenomena were explained in terms of them. To-day they are

fallen so low that a man can take a good degree in physics without really knowing what a vortex is. I have tried to show you this evening that they offer an interesting study. It may seem a far cry from the bath-room plug to the sensitive flame, but I have tried to show you that these two phenomena, as well as the resistance offered to the motion of aeroplanes, the waterspout, and many other apparently disconnected phenomena may all be appropriately considered under the heading of vortex motion.

[E. N. DA C. A.]

## Friday, December 18, 1936

SIR GEORGE SIMPSON, K.C.B., C.B.E., D.Sc.
*Manager and Vice-President, in the Chair*

THE LORD RAYLEIGH
M.A., Sc.D., LL.D., F.R.S., M.R.I.

## OPTICAL CONTACT

I HAVE here a glass wedge of two degrees angle ; it is, in fact, a spectacle prism, used to make the images unite in cases of double vision. Such prisms are very useful in many of the experiments which I hope to show you. I use this prism as a reflector to reflect on to the screen the image of an electric arc. If the two faces were parallel the reflexions would overlap, but as they are inclined the images are seen separate and about equally bright. Now I superpose a second prism on the first, and you see that there are three images (Fig. 1). The two faces which are together reflect light, just as they did when they were separate, and in fact the central one of the three images which comes from this compound face, or interface as I may call it, is brighter than the others.

This is what generally happens when two glass surfaces are put together without special precautions. They do not come into close contact, and their separate individuality is completely preserved. Yet this is not what might be expected. If the two glasses were really close, they would be equivalent to one piece, and no reflexion would be observable in the interior. Ordinarily this does not happen, because various agencies prevent the glasses coming close enough together. Thus par-

ticles of dust and traces of grease tend to do this, and, moreover, ordinary surfaces, such as those of spectacle prisms, are not flat enough to come into close contact, except over a very limited area.

However, by using really flat surfaces, scrupulously clean, and by sliding them together so as to exclude dust, two surfaces may be made to *contact*,
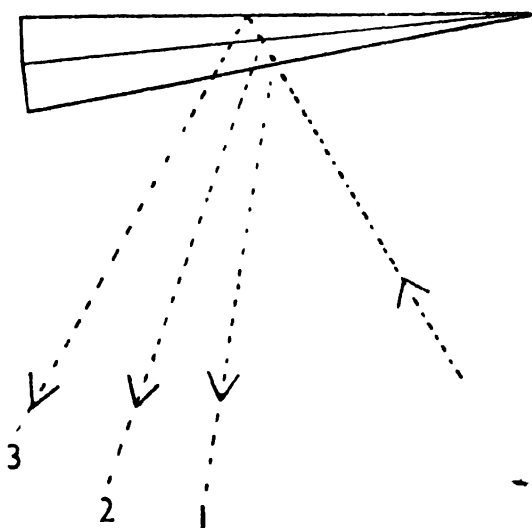


FIG .1

that is to say, they will come into contact so close that the reflexion at the interface practically vanishes, and when once this has been achieved (and it is easier said than done) the glasses adhere with very considerable force. Such glasses are said to be in optical contact.

Although, as we shall see, the germs of this subject are to be discovered in much earlier work, optical contact has mainly come into notice in connection with the manufactures of Messrs. Adam

Hilger, Ltd., under the management of Mr. F. Twyman, F.R.S. Mr. Green, the foreman of their optical shop, has been prominent in this. To mention one instance, the echelon transmission gratings made by the firm have the glass plates put into optical contact, which improves the transparency and makes the instrument much more efficient. I have heard the late Professor Michelson, who was perhaps not very lavish of praise, speak with enthusiasm of this great improvement in his invention.
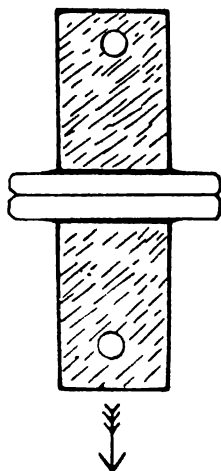
You remember the three images from the two superposed spectacle prisms, the central image being the brightest. I have here two prisms which are similar to those used before, except that the faces are very flat. The two adjacent faces have been contacted. I throw the reflexion on the screen, and you will be inclined to say that there are only two images, those from the outer faces, the interface reflexion having disappeared. But look again, I think you will see that there is a very faint image halfway between the other two. So that the interface, even of contacted glass, does give a faint reflexion. I shall have more to say about this presently.

I want now to show you the strong force of adhesion between two contacted glasses. The glasses which I propose to use have an area of less than 2 square inches. They have brass blocks cemented on to the backs and cords are attached to

56 lbs

Fig 2

these blocks. You will see that by pulling on the upper glass we can lift a 56 lb. weight, merely by the adhesion of the glasses (Fig. 2). On occasion contacted glasses have been made to stand even considerably stronger pulls than this, up to 56 kilograms per square centimetre. However, there does not seem to be anything very definite about it, and there is little doubt that the separation starts at points of local weakness.

I have mentioned that the germ of this subject can be discovered in the writings of long ago. Let me refresh your recollection of the well-known experiment of Newton's rings. I have here two glass wedges, similar to those we have used in earlier experiments, except that one of them has
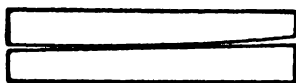


FIG. 3

a convex face of 50 ft. radius (Fig. 3). The wedges are clamped together, without any special precautions, except ordinary cleaning, and we get a circular patch of optical contact, the familiar black centre which in this case is actually about 5 mm. diameter. You see it on the screen, surrounded by a system of coloured rings. In this and other kindred experiments which I shall show you, precautions are taken to prevent the light from the outer faces of the (opposed) wedges, reaching the screen and diluting the colour effects. A few inches beyond the objective lens the arc comes to a focus, or rather to two foci, one reflected from the interface, and the other reflected from the outer faces, parallel to one another, but inclined to the interface. The latter image of the arc is stopped off, thus purifying the interface reflexion.

It should not be necessary at the present time

to stop to explain the colour of Newton's rings and their origin in the interference of light. I must take that for granted. We are more concerned here with the black centre. It is interesting, however, to recall that the blackness of the centre was felt to be a difficulty by the early exponents of the wave theory of light. Yet a simple consideration shows that it ought to be black, the consideration namely that an infinitely thin crack is equivalent to no crack at all, and therefore cannot reflect light.

We obtain local optical contact at the centre of Newton's rings without the technical difficulties which are encountered in contacting flat plates. This is natural, because all the pressure that we use is taken locally, and if there is any dirt or obstruction in the way it is crushed.

It may occur to you as a discrepancy in what I have been saying, that the optical contact which we get in this case is not accompanied by strong mechanical adhesion. This fact will conveniently serve as a starting point in considering the character of the adhesion between flat plates. Although, as we have seen, the adhesion is very strong in resisting a fair pull, it is not at all strong in resisting what I may call an *unfair* pull, that is to say, a force tending to strip the plates apart. Such a force can be applied by wedging or pulling the plates apart at one edge or one corner, and we then overcome the adhesion in detail, for the force applied is all concentrated at the edge of the contacted area. It is like tearing a strip of calico when a start has been made by making a short cut with the scissors. All the force is concentrated locally, the breadth of the strip of calico contributing nothing to increase its strength. You will see that in the case of Newton's rings, when the convex

A
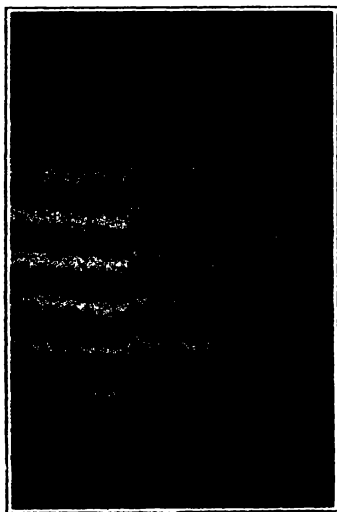


B

PLATE I

surface is squeezed flat locally, as soon as the force is relaxed, the elasticity of the material brings a stripping force of the kind to bear, and the two glasses separate themselves easily.

If, however, the convexity is very slight, and the amount of flattening involved in contact is consequently small, we may obtain Newton's rings with adhesive contact, no clamping being necessary to keep the glasses together. I show you on the screen a specimen of this kind. The dark central patch measures about 15 × 20 mm. and as you see, the surrounding rings are on a bold scale (Plate I A). This case may be regarded as intermediate between the case of Newton's rings and the case of accurately contacted flat plates.

Newton's rings result when we have contact in the middle and separation outside, but the converse case may also occur. If a piece of glass gets scratched, experience shows that it becomes difficult to contact it. The reason of this is that the hard particle that causes the scratch, ploughs up a sort of ridge, which of course prevents contact near it. If the scratch is short, we may nevertheless obtain contact at a distance from it by the pressure of the glass. I will show you on the screen a specimen of this kind. You see the scratch in the middle, the interferential tints all round, with contact and blackness at the circumference. If it is allowable to be a little Irish, we might call this specimen " Newton's rings with the black centre at the outside." The interferential tints enable us to determine the height of the ridge. In this case the increased optical distance is two waves. Remembering that in reflexion the actual distance counts twice over we conclude that the ridge is 1 wave ( $\frac{1}{40000}$ inch ) high.

Although to a first approximation the reflexion from the interface of contacted plates is black, or in other words non-existent, yet as we saw in an experiment already shown, there is a slight reflexion. In the ideal case when the surfaces are in contact the reflexion should be evanescent, but the intensity should increase as they are separated, all this occurring within the range when the inter-face reflexion is still black to the casual observer. The wave theory allows us to connect the intensity of reflexion with the distance, and it was hoped in this way to obtain a direct measure of the distance between the surfaces, which might be brought into relation with the energy required to separate them.

I found, however, that there was a very curious uncertainty in the amount of reflexion, particularly when surfaces of silica glass were used. I mean that one pair of contacted surfaces would give quite a different reflecting power from another, sometimes ten times as much. What could be the explana-tion ? It was natural to seek it in the character of the surfaces themselves. Since it is not so easy to modify the surfaces experimentally while preserv-ing the flatness necessary for contacting, I thought of trying a single surface immersed in liquid of the same index, which automatically adjusts itself to the shape of the immersed surface. This at once gave a clue to the mystery. Carbon tetrachloride has nearly the same index as fused silica, and by adding a little ether an exact adjustment can be made for green, the brightest part of the spectrum, so that green light passing through a prism of silica immersed in the liquid suffers no deviation. In this way we get a liquid which matches optically the inside of the silica prism, irrespective of what

the optical quality of the skin may be. The skin, on the other hand, determines the reflecting power of the immersed silica. If the skin is the same as the inside, then there should be no reflexion. When I tested a surface commercially polished, there was usually very little reflexion, and if a fired surface, or a surface slightly washed with dilute hydrofluoric acid was used, the reflexion became practically evanescent. On the other hand, any attempt to polish it further with the somewhat amateurish technique of my laboratory increased the reflexion largely ; and rubbing it on a dry felt wheel with a very fine abrasive powder did so still more, till a practical maximum had been reached. An abrasive powder used in this way does not destroy the polish. The increase of reflexion produced is so great that it can readily be seen even without immersion. I have here a piece of silica, half of which has been rubbed on the wheel, and the other half left with the commercial polished surface untreated. To eliminate reflexion the back is blacked. You see on the screen how definite is the increased brightness resulting from the treatment.

If we annul most of the reflexion by immersion, there is of course much less light, though much more contrast. The deficiency of light is rather unfavourable for a projection experiment. However, I project a strip of silica treated on the felt wheel, on which a cross has been etched with dilute hydrofluoric acid, so as to expose a surface of untreated material. You see that this cross shows dark on a bright ground.

The facts which I have put before you show that the felt wheel with the fine abrasive produces a modified skin ; and it is interesting to enquire what is the optical character of the skin and, also,

what is its thickness ? By putting the surface in a series of refracting liquids, we can find which liquid reduces the reflexion to a minimum. I find that a mixture of 75 volumes of carbon disulphide and 25 volumes of benzene, index 1.596, does this. It is to be concluded that the skin is quite as refractive as light flint glass, whereas the body of the silica (index 1.46) is much less refractive than crown glass. This is a surprising result, and shows that the skin produced by the treatment is something quite different from any of the known modifications of silica. Yet it is difficult to see what else than silica can enter into its composition. By dipping part of the plate in dilute hydrofluoric acid we may dissolve off the modified layer from that part, leaving a kind of step. To estimate the height of this step, we apply a flat test plate and observe the dislocation of interference fringes when they cross the step. In this way the height of the step, and hence the thickness of the layer, has been estimated at $\frac{1}{50}$th of the wave length of light in air. Plate I B shows a strip of this kind ; it has, however, been etched more deeply than is necessary to remove the layer. Similar effects to those with silica can be obtained with glass surfaces, but the layer produced by the felt wheel is definitely less refractive than in the case of silica, and since the material is more refractive to begin with, the change is a good deal less conspicuous.

Returning to the contacted surfaces, it will now not be difficult to understand why some have a stronger reflexion from the interface than others. It is because the ordinary process of polishing is somewhat haphazard. It may produce a modified layer, as the felt wheel with fine abrasive does, or it may remove the layer already produced. This

depends on the exact conditions of the polisher. I have verified this, using a cloth polishing wheel with rouge, which was quite wet to begin with and was allowed to dry. The reflexion of the silica surface was tested at intervals, and was found to pass through a minimum at a certain degree of moisture of the polisher. Concurrent weighings showed that this was the stage at which material was being most rapidly removed.

It is evident that if the contacted surfaces are not exactly alike, residual reflexion is to be expected, even if there is no separation of the surfaces. Further, even if they are alike, there will be reflexion at the separating surface between the skin and the body of the material in each of the glasses. If therefore we wish to reduce the interface reflexion to a minimum, we must get rid of the skin. I have tried contacting two glasses which have been deprived of their skin by means of hydrofluoric acid or caustic potash, but this is not as satisfactory as might have been hoped, because the chemical treatment develops all kinds of defects which were latent in the polished surface, and the light scattered by these makes it very difficult to investigate the faint residual reflexion. I have not been able to determine with certainty that the reflexion is in any case less than would correspond to about 10 Angstrom units distance between the planes. This is three or four times as much as the distance apart of the atoms in crystals.

In conclusion I wish to explain how we may determine the amount of work done in separating two contacted surfaces, or even the amount of work which can be gained by the action of the attractive forces when they are contacted. These amounts of work are not the same, for the process

of decontacting and contacting again is not, in the thermodynamic sense, reversible. We must start with a contacted surface and slowly tear the contact away. The last experiment I shall show you will illustrate how this may be done. We have a square of glass 3 mm. thick, contacted diagonally on to a larger piece, but one corner of the square overlaps the edge (Fig. 4). For security the two pieces are clamped together as shown by the dotted outline, and the non-contacted surface of the smaller glass is blackened to diminish false light.
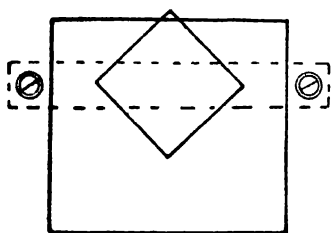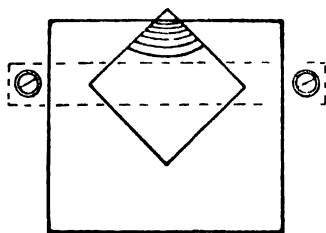


FIG. 4                FIG. 5

I project the reflexion through the larger basal glass. Initially all is dark on the screen. I pull the corner of the smaller glass away from the larger one with my forefinger, the contact is stripped away and interference fringes spring into view (Fig. 5). These extend as the pull is increased, five or six orders of colours being brilliantly developed. On releasing the corner, the glass becomes flat again, the contact is restored, the fringe system recedes, and a moment later it vanishes entirely, leaving all dark as at first.

The quantitative experiment is on the same principle, but the pulling apart is done by weights gradually added, instead of by the finger. The glasses are horizontal, and to find how far the

weights attached to the corner have descended, we count the number of interference fringes. This shows how much the corner, initially in the plane of the larger glass, has been depressed from it. Hence the work of decontacting may readily be calculated. On removing the weights, the plate recontacts, and the work recovered may be determined. The work spent in bending the glass is readily allowed for. Experiments made in this way gave 71.2 ergs. per square centimetre for the work of stripping, and for that recovered by recontacting about half this amount.

It is interesting to note that the work done in creating a liquid surface (surface tension x length) is of the same order as this.

[R.]

## RECENT CRYSTALLOGRAPHY

BENEATH the immense variety of Nature's structures lie certain remarkable simplicities. Thus the elements used in the building are surprisingly few in number. There are but ninety-two in all, of which only a few are of common occurrence, and many are used very sparingly. Indeed there are one or two of which it can only be said that they should exist, though as yet they have not been found. Oxygen constitutes half of the known world, half the remainder is silicon ; aluminium comes next. In living things carbon is one of the most important elements, yet it constitutes only a fraction of one per cent. of the world as a whole.

It is becoming clear that Nature is most economical in her designs also This very remarkable phenomenon is especially evident in the structures of living matter. The proteins which play such a predominant part in the animal are all based upon a certain fundamental pattern, and so is the cellulose which plays a corresponding part in plant-life. The recent discoveries of the biochemists have revealed the existence of several classes of substances, as for example the sterols which though minute in amount have a powerful effect on life and health. In each class a basic pattern is common

to all, and the rich variety is derived from a common theme.

There is a further simplicity in the regularity of arrangement of the atoms and molecules in all substances, or at least in the effort to achieve regularity. It is naturally most in evidence in the solid body. When the solid forms from the melt, or assembles out of solution, or grows from deposited vapour, the atoms and molecules of which it is made settle themselves in orderly array. The process is often sensitive to disturbance of the surrounding conditions, but if it is allowed to go on as it begins, it ends in the production of a visible crystal. The constancy of the angles which the faces of a crystal make with one another is evidence of the regular repetition in space of some fundamental collection which is the unit of pattern of the whole. Crystallographers have long surmised that the crystal is the consequence of such a construction.

The X-ray methods of analysis have made it possible to observe and measure these regular details of the assembly. The methods have often been described ; it is sufficient to say now that they depend on the reactions between the fine and regular succession of the X-ray waves and the fine regularity of the crystalline arrangement, the two degrees of fineness being fortunately of the same order. The study of crystalline matter with the help of X-rays is one of the most powerful methods of investigating natural structures. It has contributed materially to the realisation of the fundamental simplicities in the choice of design.

When the X-ray methods were introduced they were first applied to the determination of the simpler crystalline structures—rock salt, diamond and the like. Since then the improvements in tech-

nique and in theory have been so great that complicated structures have been completely worked out, such as for example the silicates and some of the less complex organic crystals ; and it has been possible to go some way in the examination of the complicated and large molecules which are found in living organisms, such as the proteins to which reference has already been made.
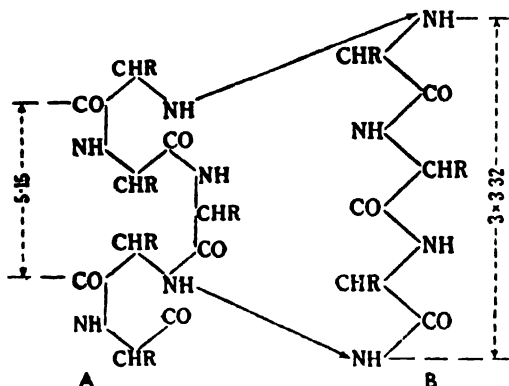


FIG 1.—A. A keratin chain partly coiled up Notice the approximation at certain points of the NH and the CO.   B  The extended keratin chain. (*Astbury*).

The first of the proteins to be examined in detail was ' keratin,' the important constituent of wool, hair and horn.  It is a remarkable illustration of the persistent occurrence of regularity in the structures of Nature, and at the same time of the power of X-rays to discover it, that crystalline structure should have been proved to be an important feature of substances which are ordinarily supposed to be anything but crystalline.

The fundamental feature of keratin is the constant repetition in a zigzag line of the atoms carbon, carbon, nitrogen in long succession.  The same

repetition is found in all the proteins. These atoms carry certain attachments, some of them invariable, others not. In the protein chain when extended to its full length, each nitrogen carries a hydrogen atom, and one of the two carbons in each unit carries an oxygen to which it is joined by a double bond. The other carbon carries a hydrogen, and is also attached by a single bond to one end of a secondary or side chain, which varies from protein to protein, and generally from portion to portion of the same protein. In the accompanying diagram (Fig. 1), taken from Astbury's " Fundamentals of Fibre Structure," the side chains are denoted by R, R, . . . The difference between one protein and another so far as constitution is concerned is due to differences in the nature of the side chains.

By X-ray measurements it has been possible to determine the length of the repeat in various forms of keratin, and to show its constancy, and its agreement with the suggested structure and with the values of interatomic distances as found in other organic crystals.

Svedberg showed in 1929 that by means of the greatly accelerated settling of a suspension subjected to centrifugal force, it was possible to estimate the weights of various proteins. He found that nearly all of those which he examined had a molecular weight of about 35.000, or a simple multiple of this number. In blood he found certain others of still greater molecular weights, to be counted in millions. The constitution of all proteins is about the same, that is to say, carbon 50–55%, hydrogen about 7%, nitrogen 15–19%, oxygen 19–24%, with traces of sulphur, phosphorus and metals. In keratin the sulphur content may amount to 2 or 3%.

Proteins are assembled in the course of plant-life. Animals require proteins for their nourishment, but do not themselves manufacture them to any considerable extent. They must therefore eat plants, or eat animals that eat plants. Emil Fischei showed, at the beginning of this century, that the assemblage of a protein was accomplished by the successive linking of amino-acids into a long chain, water molecules being shed in the process. The
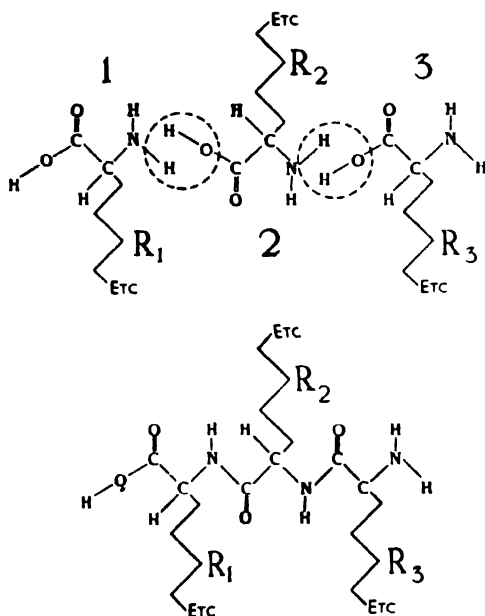


FIG. 2.—In the upper portion of this figure three amino-acid molecules are drawn side by side, as if they were ready to be linked together. In the lower portion the junctions have been made: with the shedding of water molecules formed by the combinations of the atoms within the circles of the upper portion. The side chains R may vary in form. If they are, alternately, a single hydrogen atom and a methyl group (CH₃) the particular protein is the constituent of natural silk.

amino-acids are long-chain molecules, at one end
of which (and sometimes at both ends) are found
the basic group $NH_2$ and the acid group COOH.
The linkage is effected by the interaction of the
basic group of one amino-acid with the acid group
of the next as illustrated in Fig. 2.

It is an accepted hypothesis that two parts of a
molecule, separated from each other by a single
bond, can be rotated with respect to each other
about that bond. The long chain illustrated in
Fig. 1B can therefore be crumpled up into less
extended forms by successive rotations about its
bonds. Astbury supposes that the keratin chains of
a wool fibre in its unstretched state are shortened
into the form of Fig. 1A. It is interesting to
observe that a nitrogen is now brought near to a
carbon which is five links away along the chain
(Fig. 1B). Frank has pointed out that a junction
must be effected here between the N and the C ;
the H belonging to the N passes over to the $C=O$
group, converting it into C–O–H. Groups of hexa-
gonal rings are thus formed, as the figure shows.

The side chains of one long chain molecule can
interact with the side chains of another, and it is
in this way that the long molecules are tied together
into a firm substance like hair or horn. When the
chain is coiled up these side chain attractions are
more and more satisfied by interactions in the
molecule itself. It is possible that in this way, as
has been particularly emphasised by Wrinch,*
something like a ball is formed, which has relatively
little attraction for other similar balls. Such may
be the case in the white of an egg in its natural
state ; boiling the egg breaks up the balls, and the
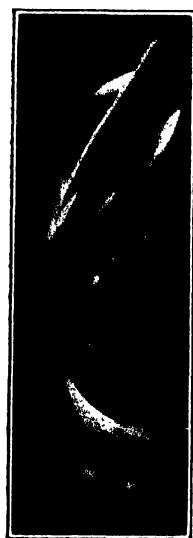uncoiling chains link together, forming the solid

*Nature Vol. CXXXVII, p.411. 1936.

mass of the white in a boiled egg. The substance is said to be ' denatured.'

These considerations, though some of them are admittedly provisional, show how the knowledge of protein structure is steadily increasing, how, in fact, the examination of the structure can be looked on as a branch of crystallography. The X-ray methods furnish measurements of distances and angles which suggest the possible forms of the proteins in space, forms which must obey certain geometrical conditions. Some of these are carried over from examinations of other organic molecules ; others are derived from direct X-ray measurements of the proteins themselves. It then becomes of great interest to connect the physical form with biological properties.

Quite recently a remarkable discovery by Dr. Stanley, of the Rockefeller Institute at Princeton, has shown that a certain virus, which causes disease in tobacco plants, is a protein. Its composition falls within the protein limits already quoted. Its molecular weight has been found by Svedberg to be of the order of 17 millions. It has reactions to polarised light which show that it is crystalline, and that the great molecule is needle-like in form. These conclusions follow from the fact that a solution of the virus shows when disturbed the effect known as ' streaming double refraction.' When minute crystals are suspended in a liquid medium they normally point in all directions. If the suspension is examined between crossed Nicol prisms, the field is dark. But if the suspension is so disturbed that the particles in any region are arranged to point in the same direction, or at least so that on the average they point more in one direction than another, the whole behaves like a single crystal, and

Hardness numbers

| | |
|---|---|
| 1A | |
| 1B | 1206 |
| 2 | 1169 |
| 3 | 1112 |
| 4 | 1132 |
| 5 | 895 |
| 6 | 290 |
| 7 Before nitriding | 280 |

C. X-ray photographs taken at different depths in a nitrided steel: the figures on the left refer to the hardness. The top figure shows that at the very surface iron nitrides have been formed.
*(By courtesy of the Director, National Phys. Lab.)*

*(Crown copyright reserved)*



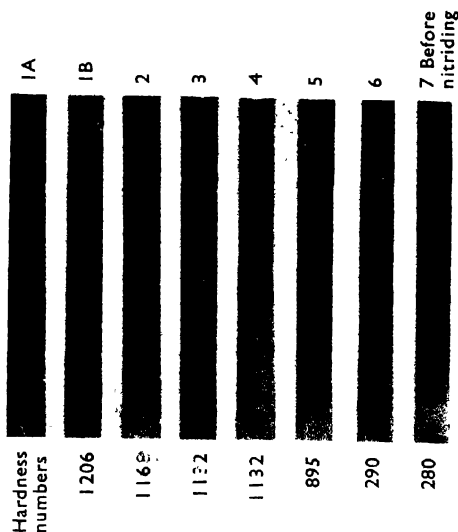A. Wake of a goldfish swimming in a dilute solution of protein from a sap infected by tobacco virus, observed between crossed Nicols.
*(Photo by Ramsey & Muspratt.) (By courtesy of J. D. Bernal)*
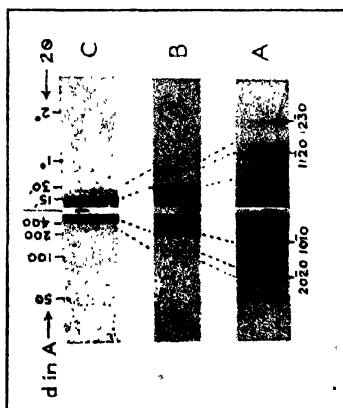


B. X-ray patterns of dry gel (A), wet gel (B), and 13 per cent solution (C).
*(By courtesy of J. D. Bernal).*

PLATE I.

illumination appears in the field. The very beautiful display of this effect by a suspension of the needle-like crystals of vanadium pentoxide has been known for some time.

The disturbance of the liquid by stirring of any kind produces whirlpools in which layers of the liquid slide past one another in shearing motion. Any long rigid particle lying across the general direction of flow is moved forward more slowly at one end than at the other. In consequence it tends to set itself parallel to the stream. But the conditions are unstable. If it continues its rotatory motion past the position of the parallelism it undergoes a further rotation of 180°. On the whole, however, one may expect the periods of parallelism to be much longer than those in which the particle lies across the stream. This effect can be demonstrated by allowing glycerine to move along a narrow glass tube, carrying with it short lengths of a fine fibre.

In a whirlpool, therefore, the rodlike crystalline particles tend to set themselves tangentially to the whirl. A beautiful illustration of this effect is given in Plate I A which is reproduced from a communication to 'Nature,'* by Bernal and others. A small fish is moving in a virus suspension between crossed Nicols. The movements of the tail leave vortices behind them and the crystalline needles of the virus follow the movement of the liquid in spirals approximating to circles. The field shows bright spots arranged at the corners of a cross ; because the line of movement makes an angle of 45° with the planes of the crossed Nicols at four points. (The effect was displayed during the lecture in a film prepared by Dr. James Gray.)

* Nature Vol. CXXXVIII, p.1051. 1936.

Bernal and his collaborators have been able to obtain X-ray diffraction photographs of the virus. These are shown in Plate I B. The pure virus in a suspension stronger than 2 per cent. separates into layers which are sharply distinguishable. One of these (B) is " an extremely soft gel well orientated and with much higher birefringence, 0.007, than the liquid. The outer part of this wet gel shrinks by 50 per cent. and forms a layer of higher refractive index, but lower birefringence, 0.003." These two layers give respectively the photographs A and B of Plate I B. The C photograph is obtained from a 13% solution in which the crystals are orientated by flow along a narrow tube. A similar set of lines is found in the three cases, indicating a hexagonal arrangement, such as would naturally be found in an arrangement of long rods put side by side. In A they are at a distance, rod to rod, of 152 Angstroms ; in B, 210 Angstroms. In C the distance is more vaguely defined as one would expect, and ranges from about 300 to 470 Angstroms. Besides these evidences of regularity in the side to side placing of the molecules, there is evidence in other photographs of much narrower spacings, which do not vary with the condition of the material and the arrangement of the molecules side by side. They refer obviously to regularities within the molecule. The length of the spacing is given as $3 \times 22.2 = 0.2$ Angstroms, a result in close agreement with measurements of Wyckoff and Corey on the tobacco virus and on a variety of it known as the *aucuba* virus. It is very remarkable that this virus, though a crystalline protein, appears to be capable of multiplying itself and of generating an anti-body, and to be liable to mutation.

Brief reference may be made to a few other

examples of the development of crystallography by means of the X-ray methods, supplemented now by the methods of electron diffraction.

On previous occasions the shadow picture of the electrons in a molecule of a phthalocyanine has been shown in the Royal Institution lecture room.
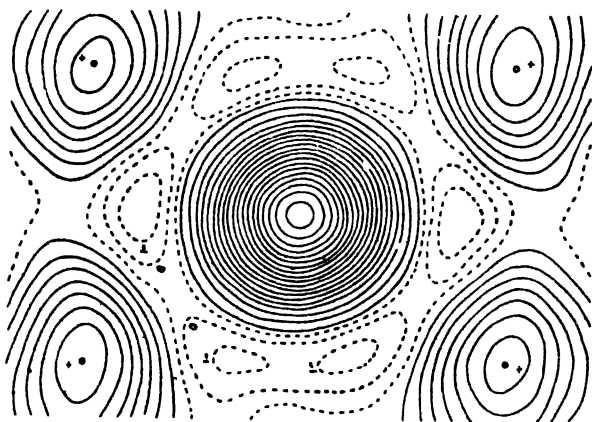


FIG. 3 —The central portion of a molecule of nickel phthalo-cyanine, showing the electron density of the nickel atom, projected on the plane of the diagram

( *J. M Robertson* )

The particular phthalocyanine was that in which the central space was occupied by two hydrogen atoms. Dr. Robertson has now extended his observations and calculations to the nickel phthalocyanine. The shadow picture contains a beautiful display of the electron distribution in the nickel atom, projected upon the plane of the diagram. The central part of the molecule containing this atom and its immediate surroundings is shown in Fig. 3. The contour lines of the nickel atom are drawn so that from line to line there is a difference of two elec-trons per square Angstrom unit. In the other

parts of the figure the difference is one electron per unit area. The accuracy is high, and the figure must give a fairly accurate representation of the distribution of the electrons in the nickel atom. It can be represented closely by an empirical formula of the form $Ae^{-kr^2}$ where r is the distance from the centre of the atom.

Applications of the new crystallographic tools to the study of practical problems of metallurgy and engineering have been numerous and important. As an example may be taken the study of the effects of cyclic stress upon a mild steel, recently described by Gough and Wood*. It is well known that metal structures can in certain cases be fractured by a sufficient application of intermittent or cyclic stresses ; that, to use the technical term, they become fatigued. Gough and Wood establish the important point that " the application of cycles of a safe range of stress (strain) is unable to cause *progressive* damage to the state of the structure." The X-ray powder photographs of a piece of mild steel show no change during the imposition of cyclic stresses of 9 tons to the square inch, up to $10^6$ repetitions. This is just within the safe range of stress. But when the stress is increased to 9.7 tons which lies outside the safe range, there is a gradual change. The X-ray photographs show that the crystallites are being broken up into powder consisting of far smaller crystal portions, and the structure fractures after about half a million cycles or so. Thus the progressive deterioration of the specimen takes place only when a certain safe load has been exceeded.

In this case the appearance of the X-ray photograph indicates the condition of the material. In a

* Proc. Roy. Soc. A. Vol. CLIV, p.510. 1936.

number of instances the photographs are now used in this way ; the spacings of the crystal planes which are determined by the positions of the lines are not in question. Thus the series of photographs of Plate I C shows the conditions of a steel at various depths below the surface when ammonia at 500° C. has been passed over it. The diffuse character of the lines shows that the crystallites have been broken up. Close to the surface the fragments are so small that the material is practically amorphous. At the surface itself iron nitrides have been formed. The physical effect is a very marked hardening as indicated by the hardness numbers in the diagram. The whole effect is contained within a quarter of a millimetre from the surface. These photographs are due to Dr. Shearer, of the National Physical Laboratory.

Diffuseness in the photograph may also be caused by any distortion of the lattice, such as occurs when a specimen is severely strained. This is found to be the case, for example, in the operation of wire drawing.

One of the most remarkable developments of X-ray crystallography is the advance in our knowledge of the internal condition of an alloy. The " order-disorder " theory, as it is termed, has been especially studied in the Manchester laboratories. It has already been described by W. L. Bragg in this theatre, so that there is no reason for a further discussion, which indeed would be impossible in the time at my disposal. Lack of time also forbids any further reference to recent developments in other directions.

[W. H. B.]

# Friday, February 12, 1937

SIR ROBERT ROBERTSON, K.B.E., M.A., F.R.S.
*Treasurer and Vice-President, in the Chair*

I. M. HEILBRON, D.S.O., D.Sc., Ph.D., F.R.S.
*Sir Samuel Hall Professor of Chemistry and Director of the
Chemical Laboratories in the University of Manchester*

## PIGMENTS ASSOCIATED WITH THE FATTY TISSUES OF PLANTS AND ANIMALS

THE varied and beautiful colours elaborated by Nature make a universal æsthetic appeal, and to the chemist there is the added attraction of interpreting them in terms of molecular architecture. Despite the great diversity of hues the actual number of fundamental types is comparatively small. Thus the great classes of plant pigments are the water souble anthocyanins and anthoxanthins, and the fat soluble plastid pigments which comprise the green chlorophyll and the yellow to violet lipochromes. This latter class, which it is my intention to discuss with you this evening, comprises a sharply defined group of compounds, not only widely distributed in flowers and fruits, but also met with throughout the whole of the animal kingdom. In virtue of the fact that the first of these pigments to be obtained in crystalline form was isolated from carrots, they are also commonly known as carotenoids. As a class they are characterised further by their insolubility in water, by their sensitivity to atmospheric oxygen, which readily bleaches them, and by the characteristic

deep blue colouration which they give with sulphuric acid. In addition to its occurrence in carrots, the most typical member of the group, carotene, is invariably found with chlorophyll in all green foliage, and is also associated with another yellow constituent to which the name xanthophyll was originally given. Despite the wide distribution of these two pigments in Nature it must be emphasised that they, in common with other members of the group, occur only in very minute amounts in living matter. Thus the fresh carrot contains only about 0.1 % of carotene, while in fresh grass the order approximates only to 0.01 % carotene and 0.02 % xanthophyll.

Although carotene was originally isolated as early as 1831, the first work of real chemical significance in this field is due to Willstatter and his collaborators, who during the years 1906–1914 effected the isolation of several of the more important carotenoids in crystalline form, and established their empirical formulæ. It was thus found that carotene was a hydrocarbon having the formula $C_{40}H_{56}$ and that it was isomeric but not identical with lycopene, the pigment of the ripe tomato. Similarly leaf xanthophyll was shown to have the formula $C_{40}H_{56}O_2$ and to be very similar to, if not identical with, the yellow pigment of egg yolk, then known as lutein. Another interesting member characterised at this period was a still more oxygenated pigment, fucoxanthin, $C_{40}H_{56}O_6$, which was obtained from *Fucus vesiculosus*, a member of the brown seaweeds.

The further development of the chemistry of the carotenoids languished until 1928, when Zechmeister, Cholnoky and Vrabely proved, by means of quantitative catalytic hydrogenation, that caro-

tene contained eleven ethenoid linkages giving rise to a saturated perhydrocarotene of formula $C_{56}H_{78}$, from which it can be deduced that the pigment itself must contain two cyclic systems. In a similar manner Karrer and Widmer ascertained that lycopene absorbed thirteen molecules of hydrogen yielding a paraffin hydrocarbon $C_{40}H_{82}$, a result which shows that the pigment must be acyclic.

The next important development in our knowledge of the carotenoids came about when Kuhn, Lederer and Winterstein reintroduced the method of chromatographic analysis as a means of separating and purifying the individual pigments of plant tissues. As early as 1906 the Russian botanist Tswett had shown that substances such as calcium carbonate or powdered sucrose, when packed in vertical columns, had the power of adsorbing pigments from their solutions in non-polar solvents such as light petroleum, and that if the original material were a mixture, the constituents would orient themselves as coloured zones or bands at various positions on the column, depending upon the strength of adsorption of the individual components. On applying this technique to ordinary carrot carotene it was found that the pigment could be separated into two distinct components, the optically inactive $\beta$-carotene and the dextro-rotatory $\alpha$-carotene. The former is adsorbed above the latter when the chromatogram is developed by washing with light petroleum. Of these two isomerides $\beta$-carotene generally predominates, although in palm oil-carotene there is as much as 30% of the $\alpha$-isomer. Shortly after this discovery the presence of a third isomer, $\gamma$-carotene, was discovered under conditions which serve to emphasize the elegance of this technique, which more

than any other has contributed to the recent rapid progress in this field. Commencing with 35 grams of crystalline carotene (equal to about 6 cwts. of raw carrots) Kuhn and Brockmann subjected it to chromatographic adsorption and observed a small pigmented zone above that of the $\beta$-carotene in the column. After many adsorptions they ultimately succeeded in isolating from this zone 34 mg. of the $\gamma$-isomeride, with which amount they were able to study its physical, chemical and biological properties, and so advance a structural formula for the compound, a remarkable example of the delicacy of modern chemical technique.

A further interesting and important point in connection with chromatography is that not only does it provide us with a readily accessible method for the isolation of minute amounts of specific pigments, but it provides in addition a means of differentiating·various classes. Thus the hydroxylated carotenoids (or xanthophylls as they are now generally called) are all much more strongly adsorbed than the hydrocarbon pigments, and further the degree of adsorbability is proportional to the number of hydroxyl groups.

A further criterion determining the position of adsorption in the column is the number and disposition of unsaturated linkages. This is illustrated by the fact that lycopene, $\gamma$-carotene, $\beta$-carotene and $\alpha$-carotene are retained on the adsorption column in descending order, which corresponds partly with the total number of ethylene linkages in their molecules and partly with the length of the conjugated system. This latter effect is evidenced in the $\beta$- and $\alpha$-carotenes, both of which contain eleven double bonds ; in the former, however, these are all conjugated, whereas in $\alpha$-carotene

only ten are so arranged. Table I (due to Winterstein) illustrates these general principles :—

TABLE I

THE RELATIVE ADSORBABILITY OF VARIOUS CAROTENOIDS FROM LIGHT PETROLUEM SOLUTION.

| Strongest adsorbed | Fucoxanthin | $C_{40}H_{56}O_6$ | | |
| | Violaxanthin | $C_{40}H_{56}O_4$ | | |
| | Taraxanthin | $C_{40}H_{56}O_4$ | Alcohols | CaCO₃ |
| | Flavoxanthin | $C_{40}H_{56}O_3$ | | |
| | Zeaxanthin | $C_{40}H_{56}O_2$ | | |
| | Lutein | $C_{40}H_{56}O_2$ | | |
| | Rhodoxanthin | $C_{40}H_{56}O_2$ | Ketone | |
| | Physalien | $C_{72}H_{116}O_4$ | Esters | |
| | Helenien | $C_{72}H_{116}O_4$ | | |
| | Lycopene | $C_{40}H_{56}$ | | |
| | γ-Carotene | $C_{40}H_{56}$ | Hydrocarbons | Al₂O₃ |
| Weakest adsorbed | β-Carotene | $C_{40}H_{56}$ | | |
| | α-Carotene | $C_{40}H_{56}$ | | |

The separation by chromatographic methods alone of complex mixtures of pigments, such as are often met with in Nature, is an extremely complicated procedure. Fortunately, however, it can frequently be simplified by carrying out a preliminary partition between two immiscible solvents, such as light petroleum and 90% methyl alcohol, when it is found that hydrocarbon pigments and esters pass into the petroleum layer (epiphasic pigments) whilst those pigments containing free hydroxyl groups pass preferentially into the alcohol (hypophasic pigments).

Another physical property of great diagnostic value in the study of carotenoids is that of light absorption. It is a well established principle that with an increasing number of conjugated unsatur-

ated linkages in a molecule, the absorption maxima become progressively displaced towards longer wave-lengths. This is exceptionally well exemplified among the carotenoids, which are characterised by well-defined triplet absorption bands, and the closest possible correlation exists between the position of the bands and the number and character of the conjugated centres. This is exemplified in Table II.

TABLE II.

EXAMPLES OF THE EFFECT OF UNSATURATION
ON THE ABSORPTION SPECTRA OF
TYPICAL CAROTENOIDS.

| Carotenoid | Number of Double Bonds. | | Absorption maxima. in $CS_2$ ($\mu$). | |
|---|---|---|---|---|
| | Con-jugated. | Uncon-jugated. | | |
| Rhodoxanthin | $\begin{cases} 14 \\ (2C=O) \end{cases}$ | o | 564 | 525 |
| Lycopene | 11 | 2 | 548 | 507 |
| $\gamma$-Carotene | 11 | 1 | 533 | 496 |
| $\beta$-Carotene | 11 | o | 518 | 485 |
| $\alpha$-Carotene | 10 | 1 | 508 | 477 |
| Violaxanthin | 9 | 1 | 501 | 469 |

It is beyond the scope of this discourse to attempt to describe in detail the elegant investigations of Karrer, Kuhn, Zechmeister and their respective collaborators, which have culminated in the elucidation of the constitution of these pigments. The formulæ for the four hydrocarbon pigments are given in Fig. 1, and these represent the fundamental structures upon which a large number of naturally occurring oxygenated derivatives are built.

Lycopene

γ-Carotene
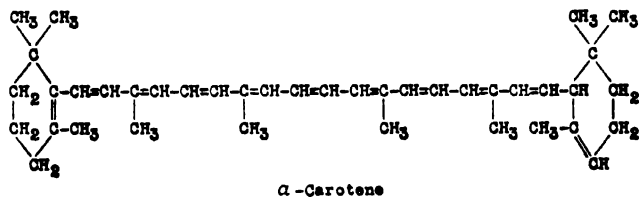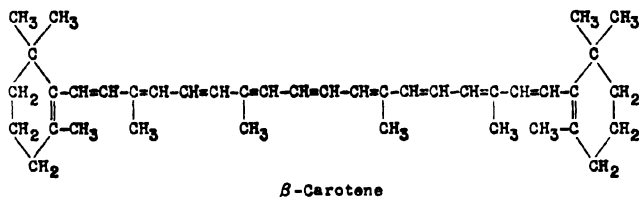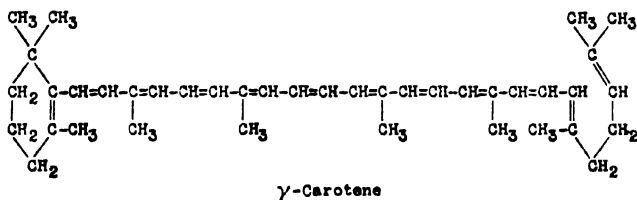
β-Carotene
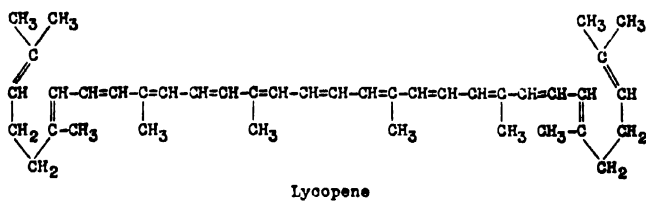
α-Carotene

Fig. 1.

In Table III the hydrocarbons and the simpler xanthophylls which correspond to them are enumerated.

### TABLE III

THE RELATION BETWEEN THE HYDROCARBON CAROTENOIDS AND THE SIMPLER XANTHOPHYLLS.

| Hydrocarbon. | Monohydroxy Derivative | Dihydroxy Derivative |
|---|---|---|
| $C_{40}H_{56}$ | $C_{40}H_{56}O$ | $C_{40}H_{56}O_2$ |
| Lycopene | Lycoxanthin | Lycophyll |
| $\gamma$-Carotene | Rubixanthin | — |
| $\beta$-Carotene | Kryptoxanthin | Zeaxanthin |
| $\alpha$-Carotene | — | Lutein |

Of the various possible types of oxygenated carotenoids, alcohols are of most frequent occurrence, and of these it will suffice to mention a few of the more common. Leaf xanthophyll, which consists mainly of a pigment having the formula $C_{40}H_{56}O_2$, has been re-named lutein and is actually a dihydroxy derivative of $\alpha$-carotene, with which it is spectographically identical. The pigment of egg yolk, originally called lutein, has now been separated into lutein proper (leaf xanthophyll) and an isomeric pigment, zeaxanthin. This is a dihydroxy $\beta$-carotene and is found in small quantity in many species of leaves, berries and fruits, but is most readily isolated from maize. The aliphatic analogue lycophyll has recently been isolated from Bittersweet. Of other xanthophylls mention may be made of flavoxanthin $C_{40}H_{56}O_3$, the yellow buttercup pigment, violaxanthin $C_{40}H_{56}O_4$, the yellow pansy pigment, and the isomeric taraxanthin from the dandelion.

The pigments so far considered have all been representatives of the higher forms of plant life, but among the Cryptogams lipochrome pigments also abound, especially among the Algæ. Recent work in this field carried out in my laboratory has brought to light some interesting generalisations. It would seem that whereas carotene is common among all algal orders exactly as in the higher plants, a definite differentiation of pigment is met with in the various species of algæ. As examples of this the Phæophyceæ are all characterised by the presence of fucoxanthin, a pigment which is very easily distinguished by the beautiful colour which it gives with 25% hydrochloric acid. Again among the Cyanophyceæ a specific pigment, myxoxanthin, is seemingly common to all classes of the group.

As already mentioned, the carotenoids are by no means confined to the plant world but are met with throughout the animal kingdom from the lowest types of protozoa to man himself. Among the vertebrates the carotenoids are absorbed from the food and consist therefore mainly of carotene and lutein. Zechmeister and his co-workers have recently made a close study of this subject and have shown that certain interesting differences are to be observed among the higher animals ; thus although the carotene-lutein ratio in grass is of the order 1 : 2, both the cow and the horse preferentially absorb carotene to the almost complete exclusion of lutein. Man, on the other hand, evinces no especial preference and both pigments are found to be present in human blood serum, while in the pig, the cat, the dog and the rodents no carotenoids appear to be retained. The early experiments of Palmer have shown that birds accumulate xanthophylls, to the almost complete exclusion of carotene, in their

blood and body-fat as well as in their egg yolk. It has also been shown by Brockmann and Völker that the yellow colour of the plumage of canaries is due to the presence of lutein, or a closely related transformation product, and is produced solely when this pigment is present in the food, for in its absence the feathers become white.

Turning now to the invertebrates, here a great diversity of carotenoids, distinct from those found in plant tissues, are met with. Among the Crustaceæ the characteristic bluish colour of the live lobster is due to the presence of a labile chromoprotein, which breaks down at the temperature of boiling water, liberating the free lipochrome, astacene. This pigment, which is a tetraketo-$\beta$-carotene, $C_{40}H_{56}O_4$, is very widely distributed in aquatic animals being found in the skin of goldfish, in starfish, the flesh of salmon, etc.

Among other marine forms the sea anemones also owe their brilliant colours to specific carotenoids of complex character ; thus *Actinia equina* contains a very beautiful purple pigment, actinoerythrin, which is an ester ; the free violerythrin, $C_{40}H_{56}O_6$ is characterised by the intense blue colour of its solution in organic solvents. Again from *Anemonia sulcata* a still more oxygenated pigment, sulcataxanthin, $C_{40}H_{56}O_8$, has been isolated.

Carotenoids have also recently been discovered among bacteria. Thus the purple rhodovibriobacteria have been shown by Karrer to contain a series of probably six closely related pigments. Of these, rhodoviolascin $C_{42}H_{60}O_2$, has been clearly characterised and constitutes the first known methylated xanthophyll.

Turning now to a brief consideration of the question of the biogenesis of the carotenoids, their

structure reveals an obvious connection with
isoprene, $CH_2{=}CH{-}\overset{\displaystyle CH_3}{C}{=}CH_2$, a hydrocarbon
that may be regarded as the fundamental unit from
which numerous important plant products are
elaborated, as a result of the regular combination
of individual units. The fact that the carotenoids
are symmetrical necessitates, however, the assump-
tion that they are built up in the plant, not on this
simple plan alone, but by the union of two mole-
cules of an intermediate compound containing
20 carbon atoms. According to Karrer this may be
the alcohol phytol, $C_{20}H_{39}OH$, which constitutes
approximately one-third of the chlorophyll mole-
cule. Condensation of two phytyl residues fol-
lowed by oxidation could thus give rise to lycopene
from which, by simple ring closure, the carotenes
would be produced.

And now in conclusion I wish to discuss the
physiological importance of these pigments in the
economy of life. As early as 1919 Steenbock
advanced evidence of a possible connection between
carotene and vitamin A, the physiologically import-
ant substance present in all liver-oils, notably those
of the cod and halibut, which is essential for
growth, for in its absence young animals cease to
grow and ultimately die. More specific effects due
to vitamin A deficiency are the eye disease, xer-
ophthalmia, decreased resistance to respiratory
infections and night-blindness. This latter affec-
tion is completely curable in a few days by the
ingestion of food rich in vitamin A ; its cause
appears to be related to the fact observed by Wald
that the vitamin plays an essential role in the visual
perception which occurs on the retina of the eye.
The vitamin itself forms an almost colourless

highly viscous oil, which is notable for its intense absorption band at 328 $\mu$ and by the characteristic violet to blue colour which it gives with a solution of antimony trichloride. It is thus clear that the pigment carotene cannot itself be vitamin A, but, as established by von Euler in 1929, it can nevertheless replace the typical vitamin in the animal organism. Complete clarification of this apparent anomaly was reached when Moore conclusively demonstrated that when carotene is fed to rats suffering from vitamin A deficiency it is converted into the colourless vitamin, which appears in the liver. The structure of vitamin A has been established, and its formation from carotene may be represented by the following equation, although the exact mechanism of conversion is still obscure :—

$$C_{40}H_{56} + 2H_2O = 2C_{20}H_{30}O$$
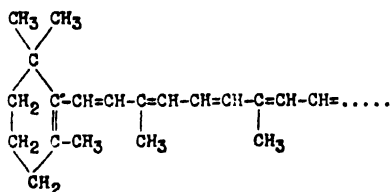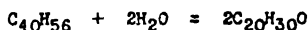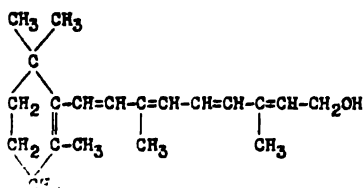


½ $\beta$-carotene



Vitamin A

Arising from the foregoing it follows that animals are primarily dependent upon the carotene of grass for their supply of vitamin A. The full appreciation of this fact opens up interesting problems for mankind, and much experimental work is being carried out in order to make use of this new knowledge.

Of the many studies which have been made of the variations in the carotene and vitamin A content of cows' milk as affected by the diet, breed, age, etc., of the cows, I need only mention a few, carried out notably by Steenbock in America, by Imperial Chemical Industries at their experimental farm in Berkshire, as well as in my own laboratory.

Palmer and Eccles were the first to prove that the yellow colour of butter is mainly due to carotene. The further discovery that butter also contains vitamin A and that the two pigments can be quantitatively evaluated by spectroscopic methods, opened the way to the investigation of the factors controlling the quantities of these substances in milk and butter. The actual amounts normally present in milk fat are extremely small—i.e., carotene 1-20 parts and vitamin A 2-20 parts per million. In collaboration with Professor J. C. Drummond of University College, London, and Dr. S. J. Watson, of the I.C.I. Agricultural Research Station, during the past few years Mr. Gillam and I have been examining the possibility of maintaining the high vitamin content of summer milk throughout the winter. We have ascertained that by using grass, artificially dried in specially designed driers, in place of the hay and concentrates usually given to stall-fed cattle in the winter, both the vitamin A and carotene content of milk (and hence of butter) can be maintained at

nearly the level usually regarded as peculiar to
summer samples. In a particular experiment one
group of Shorthorn cows was fed on a normal diet
throughout a whole winter, whilst another group
had a dried grass supplement replacing part of the
concentrates. The milk fats were examined for
both vitamin A and carotene once a month and the
results, expressed graphically in Figs. 2 and 3,
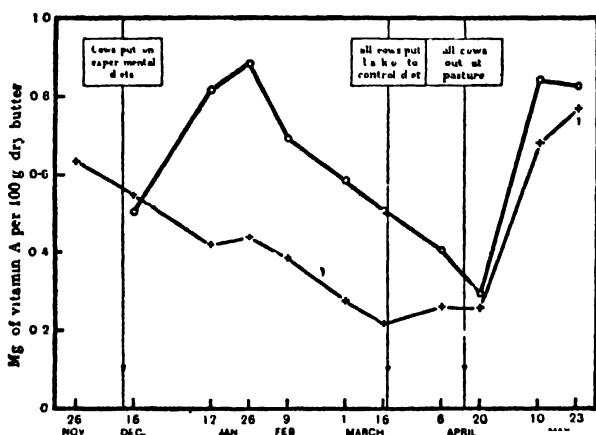


FIG. 2. Vitamin A of the butters.

leave no doubt as to the efficacy of the dried grass
supplement.

Although the vitamin A activity of a butter can
in general be roughly correlated with its yellow
colour, we have found that this is only strictly true
when comparing butters from the same breed of
cow. Thus when Shorthorn or Ayrshire cows were
fed on a ration rich in carotene, the yellow colour
rose to a maximum above which the further
addition of carotene in the diet would not increase
it. The Shorthorn butter was, however, always

somewhat more yellow than that obtained from the
Ayrshire cows. The traditionally high yellow
colour of the milk of Jersey and Guernsey cattle is
due to their having a higher " ceiling value " for
carotene in their milk-fat than any of the other
breeds. Experimental results indicate, however,
that although Shorthorn milk-fat is normally paler
than that from the Guernsey cow, the gross
vitamin A activity is about the same ; this is due
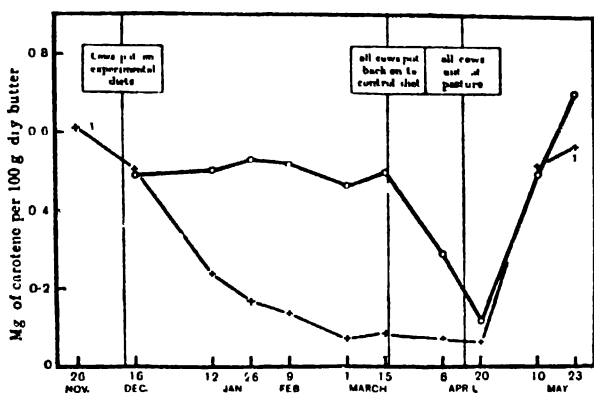


FIG. 3. Carotene of the butters.

to the fact that the milk of the Shorthorn has a
higher proportion of the true vitamin A.

The differences in carotene and vitamin A con-
tent of the milk of different species of animals are
exemplified by the Guernsey cow and the goat, for
the latter, in striking contrast to the former, gives
a butter which is almost dead white in colour.
Despite the almost total absence of carotenoids in
goat's butter, nevertheless it contains vitamin A in
only a slightly lower amount than does average
cow's butter.

Although the seasonal variations in the colour
and vitamin A content of milk and butter are

actually directly due to the presence or absence of carotene in the diet, i.e., fresh green grass is necessary to produce good yellow butter, there is one exception to this generalisation. Whatever the breed of cow or the season of the year, the colostrum is usually 10–20 times as rich in both vitamin A and carotene as is the normal milk. This quality falls rapidly from the first day after calving and reaches normality in about a week. This richness in vitamin A is therefore quite consistent with the traditional belief in the high food value of colostrum, and furnishes another remarkable example of the way in which biological processes provide for special circumstances in the life of the organism. These phenomena find an exact parallel in the human subject.

Examples of the relative vitamin A values of ordinary and colostrum milk are given in Table IV, the results having been obtained in the course of the experiments described above.

TABLE IV

COMPARISON OF THE CAROTENE AND VITAMIN A VALUES OF THE FAT OF ORDINARY AND COLOSTRUM MILK OF INDIVIDUAL COWS CHOSEN AT RANDOM.

| Breed | Date of Sample | Carotene mgms.% fat | Vitamin A mgms.% fat |
|---|---|---|---|
| Friesian   .. | Oct. 20 | 3.40 | 1.24 |
|  | Nov. 20 | 0.38 | 0.52 |
| Ayrshire..   .. | Nov. 22 | 4.60 | 2.96 |
|  | Dec. 22 | 0.18 | 0.37 |
| Guernsey   .. | Oct. 28 | 3.61 | 1.13 |
|  | Nov. 28 | 0.91 | 0.67 |
| Shorthorn   .. | Oct. 13 | 2.90 | 3.51 |
|  | Nov. 13 | 0.28 | 0.48 |

I have attempted in the foregoing to trace the stages in the development of our knowledge of the lipochrome pigments, culminating in the discovery of the close interrelation between plant carotene and vitamin A, one of the essential key-substances of animal life. How far the other members of the group have dual functions time alone will show, but assuredly the still hidden secrets of Nature will ultimately be revealed and, in the words of the late Robert Bridges, we will comprehend how

" From Universal Mind the first-born atoms draw
　Their function, whose rich chemistry the plants transmute
　To make organic life, whereon animals feed
　To fashion sight and sense and give service to man."

[I. M. H.]

A. O. RANKINE, O.B.E., D.Sc., F.R.S.
*Manager and Vice-President, in the Chair*

G. I. FINCH, M.B.E.,
D.Tech.Chem. (Zurich), F.Inst.P., M.R.I.
*Professor of Applied Physical Chemistry, Imperial College of Science and Technology*

## ELECTRON DIFFRACTION AND SURFACE STRUCTURE

THE properties of waves are very different from those of bodies in motion. For example, though two waves can pile up together on meeting under favourable circumstances, they can also extinguish each other, and this is something which it is difficult to conceive of ever happening to two colliding projectiles. Newton thought that light consisted of particles in swift motion, but when Fresnel and Young proved that light behaved like waves, Newton's corpuscular theory was abandoned. Soon after the war, however, it was found that light sometimes really did behave like a stream of particles and must therefore also be corpuscular in nature.

It seemed quite impossible to understand this duality in behaviour until de Broglie, with a typically Gallic flash of genius, boldly postulated that all particles were guided in their motion through matter by attendant wave systems. When moving through empty space the particles in their behaviour would show no signs of wave properties, but should do so when they came into contact with other particles. De Broglie calculated that the

wave-length of the waves he believed to be associated with moving particles should be inversely proportional to their mass and speed, and would therefore be difficult to detect, except in the case of particles of exceedingly small mass, because otherwise the waves would be too short to be diffracted even by natural crystals, which are the finest gratings available. But electrons are more than 1,800 times lighter than the hydrogen atom, so that the length of the de Broglie waves associated with a stream of electrons moving at, say, 50,000 miles per second should be of the same order as in X-rays and therefore detectable by diffraction by crystalline matter.

The first experimental proof of the existence of de Broglie's moving-particle waves was carried out simultaneously and independently by Davisson and Germer with slow electrons and by G. P. Thomson with fast electrons. Thomson fired a stream of electrons through a crystalline film of gold. Knowing the structure of the gold and the speed and mass of the electron, he was able to apply a quantitative test to de Broglie's law, which his experiments fully confirmed.

How are we to think of these waves which guide the electrons in their passage through the gold film ? Experiments show that they only come into action when the electrons are moving through matter; in empty space there is nothing in the behaviour of the electrons to betray their existence, and yet our experiments tell us that these waves are all pervading.

If you look at a fish swimming in water you will see that he is guided in the direction in which he moves by a wave motion which passes from his head towards his tail. When in the water the fish

B. The cross-grating type of pattern obtained when the electron beam traverses a very thin crystal



A. Electron diffraction pattern typical of the surface structure of a single crystal, in this case the facet of a diamond crystal

PLATE I

is, in a sense, like an electron passing through matter. Should the wriggling fish leap out into the air, he may then be likened to an electron speeding along through empty space, because the directions in which both fish and electrons move are then completely unaffected by their attendant wave motions. The fish model is, however, in reality inadequate ; indeed, it does not seem to me to be possible to describe de Broglie's electron waves in terms of any concrete model. Rather must we think of them as something indefinite, like waves of emotion ; where such waves are most intense, there will be the greatest chance of finding the associated bodies.

The discovery of the wave properties of moving electrons has provided us with a new and wonderfully powerful tool for the study of the structure of surfaces. Owing to their short wavelength and being electrically neutral, X-rays are very penetrating and can therefore tell us little or nothing about the structure of the surface of a body, although they reveal so much of what lies beneath. Electrons, on the other hand, carry a charge and are therefore so easily deflected by the strong localised positive charges of the atomic nuclei, that they cannot penetrate more than a few atoms deep below the surface. Thus the information afforded by diffracted electrons is virtually confined to the structure of the surface layers.

Of the several problems relating to surface structure which have recently been successfully attacked by the method of electron diffraction, one of the most interesting is that connected with the nature of polish. The late Sir George Beilby had shown, over twenty years ago, that polish does not simply consist in wearing away the projections on

a rough surface, but causes a sort of flowing of the surface material, almost as if it had been melted and then smeared like butter over the surface. Beilby concluded that the metal atoms in the final polish layer were all completely disarranged, just as they would be in a suddenly frozen liquid. Many workers have since thought that Beilby was wrong, so Thomson, hoping to decide the issue, studied the diffraction of electrons by polished metal surfaces. Now a wave probe, whether it be X-rays or a beam of swiftly moving electrons, can only give information about a body when its structure exhibits some regularly repeated feature or features. Thus, in a crystal or crystal surface, the atoms are arranged in a perfectly orderly manner which is repeated many times when the crystal is not too small. In an amorphous substance, however, the atoms are jumbled up, so that practically the only regularly repeated feature is the size of the atoms, or, in other words, their nearest distance of approach to each other. In such a case, the diffraction pattern should consist only of a few very diffuse haloes ; hence, when Thomson found that the electron diffraction pattern of polished metal surfaces consisted of broad fuzzy rings or haloes, similar to those obtained by X-rays from liquids like mercury or from amorphous substances like glass, he concluded that Beilby was right in supposing the polish layer to be amorphous. Somewhat later, however, Kirchner found that under certain circumstances metal surfaces which were known to be crystalline could also give haloes, a discovery which left the issue still in doubt.

In the meantime, while working on an entirely different problem, Dr. Quarrell and I happened to observe that a thin film of zinc crystals when freshly
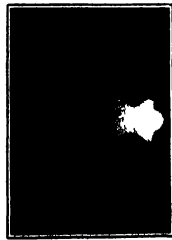
A. The halo pattern characteristic of liquids and vitreous amorphous surfaces.
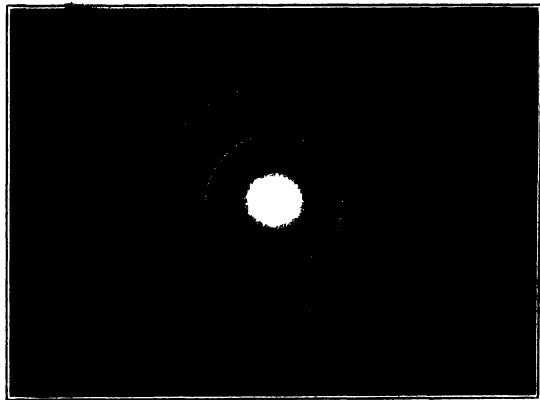


B. Pattern yielded by a thin polycrystalline film traversed by a high-speed electron beam.
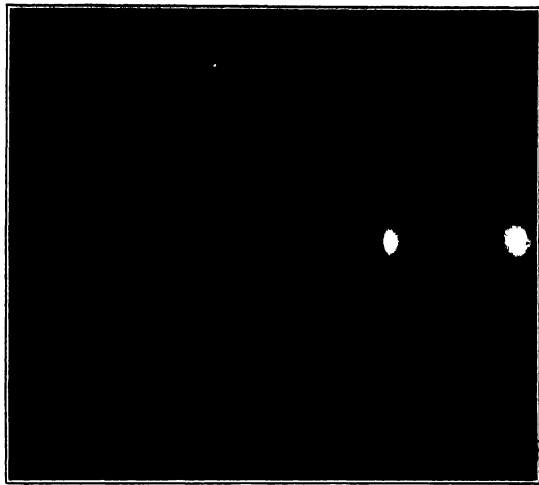


C. When crystals on a surface are orientated, i.e., "look the same way" as Newton would say, this kind of pattern is obtained.

PLATE II

deposited on a cool newly polished surface of copper gave good electron diffraction patterns characteristic of zinc crystals. The pattern, however, rapidly faded and finally disappeared, although if the copper surface had not been previously polished but was crystalline, no such fading was ever observed. The gradual weakening of the diffraction patterns meant that the zinc crystals were being destroyed and were in fact being dissolved by the polish layer of copper, which thus exhibited a property characteristic of a liquid and not shared by the corresponding crystalline surface.

How easy it must be for liquefaction of even highly refractory surfaces to occur when they are being polished has been admirably demonstrated by Dr. Bowden and Dr. Ridler. They took advantage of the thermo-electric current which flows when two different metals are joined together into a closed circuit, and one junction is hotter than the other. It is easy to measure this current and thus determine the difference in temperature between the two junctions. In this way Bowden and Ridler showed that the temperature of formation of the polish, or as it is now known the Beilby layer, on a metal surface is equal to the melting point of the metal itself.

The study of polish has important bearings upon the problem of wear in engines. The essence of the process of " running-in " an engine is a sort of vigorous polishing action, by which a deep Beilby layer of amorphous material is formed on the working surfaces. The reasons why we want to build up such a polish layer before subjecting the bearing surfaces to heavy loads are twofold. Firstly, the Beilby layer is harder and tougher than the corresponding crystalline surface.

Also, unlike polished surfaces, freshly machined surfaces are not really smooth, but have many sharp little crystalline peaks projecting above the mean level of the surface. Thus the load is not uniformly distributed over new bearing surfaces, but concentrated on to a few almost point-like areas, so that the oil film meant to prevent metal-to-metal contact is easily broken down, with the result that the high temperatures generated when the metallic surfaces rub directly against each other sometimes cause them to fuse and weld together, leading to seizure or bad scoring and excessive wear.

Although the polish layer is formed by a process of fusion and smearing of the flowed substance over the surface, the freshly formed Beilby layer on some surfaces immediately recrystallises on cessation of the polishing action. This occurs in the case of sapphire. The oxide film normally formed on an aluminium surface is amorphous, but the action of polishing causes it to crystallise into sapphire. This explains why the aluminium piston is liable to cause excessive wear of the engine cylinder. Unlike sapphire, however, spinel, a solid solution of aluminium oxide in magnesium aluminate, on polishing forms a permanently amorphous Beilby layer, which does not recrystallise like the sapphire and so produce sharp projecting crystal corners capable of cutting through the oil film to score the metal cylinder. Thus a thin coating of aluminium-magnesium alloy on an aluminium piston should lead to reduced wear, because the mixed magnesium and aluminium oxides film on its surface is "spinelised" by polishing, and the resulting spinel film remains amorphous and smooth.

[G. I. F.]

MAJOR CHARLES E. S. PHILLIPS, O.B.E., F.R.S.E.
*Secretary and Vice-President, in the Chair*

THE LORD RUTHERFORD OF NELSON,
O.M., M.A., Sc.D., F.R.S., M.R.I.
*Professor of Natural Philosophy*

# THE TRANSMUTATION OF HEAVY ELEMENTS

DURING the last few years our knowledge of the transmutation of the elements by artificial methods has grown with great rapidity, and practically all the known elements have been found capable of transmutation on a small scale when bombarded by fast particles of suitable type. By means of an ingenious apparatus called the cyclotron, Lawrence has been able to produce copious streams of protons and deuterons with energies as high as 6 million volts and moving with velocities even greater than the $\alpha$-particles from radioactive substances. Such swift deuterons are capable of producing transformations even in heavy elements like platinum and bismuth. There is some evidence that the deuteron is broken up into its constituent proton and neutron in the intense field which exists close to a nucleus. The neutron may then be captured by the nucleus while the proton escapes. For example, four radioactive elements are produced from platinum, two of which have the same chemical properties as platinum, and are thus new unstable isotopes of that element, while the other two behave like isotopes of iridium. The interpretation of the results is complicated by the

number of known isotopes of platinum, viz. masses 192, 194, 195, 196 and 198. One of the radioactive isotopes of platinum breaks up with the emission of a negative electron and the other —an unusual event for heavy elements—breaks up with the emission of a positive electron. In the first case the isotope of mass 196 is believed to be involved ; by the capture of a neutron a radioactive isotope of mass 197 is formed and this is transmuted by the emission of a negative electron into gold (mass 193). In the other case the isotope of platinum 192 forms a radioactive isotope 193, which by the emission of a positron forms a stable isotope of iridium (mass 193). One of the radioactive iridium isotopes is believed to be formed from platinum 196 by the capture of a deuteron and the emission of an $\alpha$-particle. The radioactive isotope of iridium of mass 194 is then transformed into the platinum isotope 194 by the emission of a negative electron. The origin of the other iridium isotope has not yet been settled.

The bombardment of bismuth by fast deuterons is of particular interest, as it leads to the production of a radioactive isotope of that element identical in radioactive and chemical properties with the natural radioactive body, radium E. This important result has been confirmed by showing that this artificially produced radium E gives rise to polonium—the first of the radio-active elements separated by Mme. Curie in 1897 from uranium minerals.

In general, the neutron is extraordinarily effective in producing transformations in the majority of the elements. In a number of cases very slow neutrons are far more efficient in this respect than fast ones. A suitable source of neutrons for such

experiments can be obtained by bombarding beryl-
lium with α-particles from radium. The fast
neutrons can be slowed down by allowing them to
pass through material containing hydrogen, for
example, water or paraffin. In this way more than
80 new radioactive isotopes have been discovered,
most of which break up with the emission of
β-particles. The action of neutrons on the heaviest
known element, uranium, has been the subject of
close study by Hahn and Meitner during the past
two years. Work with this element presents special
difficulties on account of its spontaneous radio-
activity. Nine new and distinctive radioactive
bodies have been observed when uranium is bom-
barded by slow or fast neutrons. All of these break
up with the emission of β-particles and with half-
periods of decay varying from 8 seconds to 3 days.
It may well be that other radioactive elements of
still longer life will yet be observed. Hahn and
Meitner have conclusively shown that not only are
three new radioactive isotopes of uranium formed,
but also radioactive elements of higher atomic
number than uranium. By the application of suit-
able chemical.methods it has been found that two
of the radioactive bodies have the chemical pro-
perties to be expected for eka-rhenium atomic
number 93, two for eka-osmium atomic number 94,
and one for eka-iridium and for eka-platinum
atomic numbers 95 and 96 respectively.

It has been found that the new radio-element
formed from uranium breaks up in a series of
successive stages analogous in many respects to
the well-known sequence of changes which occur
spontaneously in uranium and thorium. The
results indicate that three new radioactive series
are formed, two of which probably arise from the

main isotope of uranium (mass 238) after the capture of a neutron, and the third may be due to a less abundant isotope of uranium (mass 235). The two main series of transformations are believed to be isomeric and to be the consequence of two distinct varieties of transformation of the same nucleus which is formed from uranium 238 by the capture of a neutron. The possibility of such an isomeric change had been suggested some time ago in order to account for the fact that uranium X appeared to give rise to two distinct $\beta$-ray products. All the new radioactive bodies formed from uranium break up with the emission of $\beta$-particles. The active uranium isotope of half-period 8 seconds formed by the capture of a neutron is transformed successively into eka-rhenium of period 2.2 minutes, eka-osmium period 59 minutes, eka-iridium period 3 days, and eka-platinum period 2.5 days. The latter presumably forms eka-gold, but no certain evidence of transformation has been observed beyond this stage. All of these elements have nearly the same mass, 239, but, owing to the liberation of energy in the form of a $\beta$-particle, the mass of each successive element in the series must slightly decrease. With the exception of the uranium isotope of period 23 minutes, all these transmutations are produced both by fast and by slow neutrons. On the other hand, the 23-minute body can only be produced by slow neutrons of a definite energy and not by fast neutrons at all. This effect is a typical example of what is known as a resonance phenomenon of which we have many instances in other elements. The effective neutrons are strongly absorbed by uranium, and Hahn and Meitner estimate their energy to correspond to $25 \pm 10$ electron volts. From the recent work of

Dempster it is known that uranium consists of three isotopes of masses 238, 235, 234. The abundance of these three isotopes is of the order of 100, 0.3 and 0.07 respectively. It may be that the 23-minute body arises from the capture of a neutron by the isotope of mass 235. It does not appear that the bombardment of uranium by neutrons has any effect in accelerating the natural disintegration of this element. A complex series of transformations also arises when the second heaviest element, thorium, is bombarded by neutrons. Radioactive isotopes of radium, protoactinium and actinium are produced, but the exact nature of the transformations involved are still under investigation.

In the course of the lecture, the formation of $\alpha$-ray tracks was illustrated by means of a specially constructed expansion chamber. By special arrangements an image of the interior of the expansion chamber was thrown on a screen by means of the light from an arc lamp. At the moment of expansion the tracks of $\alpha$-particles radiating from a source of polonium in the middle of the chamber were made clearly visible to the whole audience. The apparatus for this purpose was kindly prepared by Dr. E. Bretscher of the Cavendish Laboratory. Experiments were also shown to illustrate the formation of radioactive bodies by exposure to slow neutrons, the elements indium and silver being used for this purpose. The marked $\beta$-ray activity produced in these elements and their rapid decay was shown by using a Geiger $\beta$-ray counter connected with a loud speaker. A preparation of uranium freed from uranium X, was exposed to a source of neutrons and the large increase of $\beta$-ray activity as a result

of a few minutes exposure was illustrated in the
same way.

[R.]

# Friday, April 23, 1937

SIR ROBERT ROBERTSON, K.B.E., M.A., F.R.S.
*Treasurer and Vice-President, in the Chair*

PROFESSOR DR. H. MARK
*Director, Chemical Laboratory of the University of Vienna*

## THE SYNTHESIS OF LARGE MOLECULES

### INTRODUCTION

THE normal molecules with which organic chemists have worked for many years, and which have led to a number of interesting and important chemical combinations, have molecular weights between 50 and 2000. Only a very few of them are smaller, and few are larger, but even then they do not exceed these limits very considerably. In the present discourse I have to speak on the synthesis of molecules built up by a much greater number of atoms and hence of a much higher molecular weight.

It has long been known that certain types of chemical reactions lead to products which do not crystallise and which are known as " resins " or " resin-like substances." Their identification and reproduction was very difficult and hence for a long time they have not been scientifically investigated. But these substances possess mechanical, thermal and electrical qualities which render them very important from the technical aspect and interest in their structure and their synthesis has recently increased.

On the other hand the study of the natural products such as cellulose, proteins, rubber, starch, silk, fibroin, chitin, etc., led to the conclusion that the

building principle* of them was the same as that of
the synthetic resins, and attention was directed to
the biochemical and technical qualities of this
group of bodies. In consequence a number of
scientists began an investigation of large molecules
and our knowledge of them increased rapidly. It
was found that two types of large molecules exist—
one with chain-like particles, another with large
molecules, having the structure of a two- or three-
dimensional network. In both cases the number of
atoms which are held together by normal chemical
main valencies is very large—between $10^3$ and $10^6$
—and hence the molecular weight lies between
$10^4$ and $10^6$ based on oxygen 16.0. With the mole-
cular weight several other qualities of these large
molecules were studied and cleared up—e.g., their
shape, their reactivity, their mechanical and
optical behaviour, etc.

When our knowledge of the outstanding quali-
ties of these bodies had reached a certain point,
chemists became especially interested in the type
of chemical reaction by which such macromole-
cules are built up, and it is the aim of the present
discourse to give a very short survey of the present
state of our knowledge of this type of chemical
reaction, the so-called polymerisation and polycon-
densation reactions.

## THE POLYMERISATION REACTION

Small organic molecules with one or more double
bonds have the peculiarity of polymerising under
certain conditions of temperature, pressure, etc.
This polymerisation leads mostly to macromole-
cular substances, which are therefore also often

* See K. H. Meyer and H. Mark : The Structure of High Poly-
meric Substances ; Leipzig, 1930 ; H. Staudinger : High Molecular
Substances ; Berlin, 1932.

called high polymers or high polymeric substances. Let us consider the most simple chemical molecule with a double bond, ethylene and its derivatives.

In this case the total polymerisation reaction (of an ethylene derivative) is given by the relation.

$$\text{n. } CH_2 = CHX \rightarrow (CH_2 . CHX)_n \qquad (1)$$

In this equation X signifies the substituent ($H$, $CH_3$, $Cl$, $C_6 H_5$, $OH$, etc.), n is a number of the order of magnitude between $10^2$ and $10^4$. The left side of the relation (1) represents a great number of independent small molecules, the structure of each of which is completely known and contains a double C—C linkage. The right side gives a large molecule, the structure of which is not quite clear in all its details, beyond the fact that it contains all atoms brought by the n molecules, linked together by main valence bonds.

Frequently the macromolecules built up from ethylene monomers are chains and in these cases one may write the equation (1) in a more detailed form.

$$\text{n.}CH_2 = CHX \rightarrow CH_3 - CHX - (CH_2 - CHX)_{n\ 2}$$
$$CH = CHX \qquad \qquad (2)$$
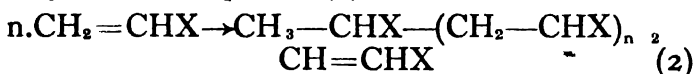
In this case a certain assumption is made as to the internal structure of the chainlike macromolecule, an assumption which is justified in some cases by considerable experimental material. Many chain polymerisation processes have been carried out with substances of this kind and formulated by relations of the type (2) and it may be pointed out that Staudinger in Germany and Carothers in U.S.A.* have worked systematically in preparing macromolecules by the aid of such processes.

But besides the body which is built up by such

* H. Staudinger, loc. cit. ;   W. Carothers, J. Amer  Chem. Soc. lui., 2527, 1931.

a process and which occupied the whole attention of the investigators, the mechanism of the process by which such long chains are formed is also of great interest and should be studied systematically. Accordingly, therefore, in the last five years we have carried out a series of experiments in Vienna to elucidate the different steps of typical polymerisation reactions.*

At once it is clear that the numerous independent particles on the left side of equation (1) cannot be linked together in one single collision process, which would have a vanishing probability, but that there must be a kind of growth which builds up the long chain. This growth process is the object of our investigations. It seems to be not only of scientific interest in connection with the physico-chemical problem of reactivity, but also of technical and biochemical importance. The technical significance is confirmed by the fact that a great number of synthetical high polymers are of considerable interest, such as artificial rubber, insulators, varnishes, etc., while the fact that our own body and the skeleton substances of all plants and animals are built up of high polymeric substances, and that therefore the growing processes of organised nature may have some connection with the studied reactions gives it biochemical importance. In any case a knowledge of the simplest polymerisation reactions will be indispensable for the understanding of natural growth. We thought, therefore, that it would be worth while to devote some work to a thorough investigation of the kinetics of polymerisation.

* H. Dostal and H. Mark: Zeit. Phys. Chem. (B), xxix. 299, 1935; Trans. Far. Soc., xxxii. 54, 1936 ; H. Mark and R. Raff : Zeit. Phys. Chem. (B), xxxi. 275, 1936 ; H. Dostal and R. Raff : Zeit. Phys. Chem. (B), xxxii. 417, 1936.

The general situation was not unfavourable ; the kinetics of normal chemical reactions, even of a rather complicated type, have been cleared up in recent years by Abel, Bodenstein, Bonhoeffer, Hinshelwood, Polanyi, Rideal, Semenoff, H. S. Taylor and others ; the experimental procedure is not too complicated and the methods of following gradually the rate of the process are more or less worked out. Besides our own studies[*] in the last four or five years, E. Bawn,[†] W. Chalmers,[‡] P. Flory,[§] Melville, Norrish, E. K. Rideal,[||] G. V. Schulz,[**] Taylor, Vernon[††] and others have contributed valuable results and one can summarise the present state of our knowledge in the following way.

The whole polymerisation process is—even when only chains are built up—very complicated and depends in a high degree upon the experimental conditions of temperature, pressure, solvent, presence of catalytically active bodies, etc. But in the simpler cases—and only these have been studied carefully up to date—one can always distinguish three typical steps in a reaction of this kind.

*a. The building of germs or nuclei.* No polymerisation can take place if there is not an initiation process by which unsaturated and highly reactive germs are formed. This process can be monomolecular, bimolecular or of another order, and furnishes the centres of growth. It consists of

* These were carried out in the Chemical Institute of the University in Vienna by Breitenbach, Dostal, Jorde, Marecek, Pilch, Raff, Rudorfer and Suess.

† E. Bawn : Trans. Far. Soc., xxxii. 178, 1936.
‡ W. Chalmers : Can. J. Res. vii. 113, 1932.
§ P. Flory : J. Amer. Chem. Soc. lix. 24, 1937.
|| E. K. Rideal : Trans. Far. Soc. xxxii. 3, 1936.
** G. V. Schulz and E. Hüsemann : Zeit. Phys. Chem. (B), xxxiv., 187, 1936.
†† H. J. Taylor and A. Vernon : J. Amer. Chem. Soc. liii. 2527, 1931.

the fact that in a certain molecule a very high amount of activation energy or unsaturation is concentrated, either by impact with another particle of high energy, by the absorption of a photon, or by the formation of an intermediate combination. In any case this germ-building reaction is rather slow and very dependent on the temperature or the presence of a catalyst (including light).

This first step has also been called the *starting process* and a thermal bimolecular start reaction can be formulated, e.g.

$$CH_2 = CHX + CH_2 = CHX \longrightarrow CH_2 - CHX - CH_2 - CHX - \quad (a)$$

where the nucleus on the right side has two free chemical main valencies, and therefore represents a very high free energy. Measurements by different investigators showed mutual agreement in the result, that the activation energy of the start reaction in the case of polymerisation of ethylene derivatives lies between 20,000 and 30,000 cal. per mole. The reaction is therefore slow and the germs are produced with a limited velocity.

*b. The reaction of growth.* If an unsaturated germ is formed in the solution there are two possibilities for its further fate. It may collide with another molecule (monomeric substance or solvent) and be de-activated by the impact. In this case the nucleus — just produced — is annihilated again and no polymerisation is started. But the germ may also react with the colliding particle and, if it is a monomeric molecule, may add it according to the relation

$$-CH_2 - CHX - CH_2 - CHX - + CH_2 = CHX \rightarrow -CH_2 \ldots CHX - (b)$$
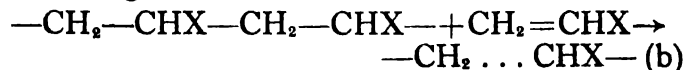
This is a step of growth and processes of this type

lead to the desired result, namely to long chains, supposing that they follow one another very quickly. Really high polymeric bodies can only be formed when a slow germ production is followed by a quick growth reaction.

In fact the rate of reaction (b) is very much higher than that of (a). Measurement shows that (b) goes at least $10^4$ times quicker; this means that a nucleus is built that begins to grow at once and rapidly gives a chain of considerable length. This high velocity is easily understood; the free valencies at the end of the germ and of all intermediate members of the growing chain, represent a high energy and therefore the activation energy of the addition reaction (b) is rather low; it lies between 4000 and 8000 cal. per mole. and explains the quick and nearly temperature-independent growth. Very interesting results relating to the sterical factor of growth would lead us much too far and must be suppressed in this discourse.

If the growing chains meet no obstacle during their period of adding monomolecular material, they will grow so long as molecules of the polymerising substance are available, and the result of the whole process would depend only on the velocity of germ-formation $k_1$ and of growth rapidity $k_2$. The competition of these two influences has been discussed quantitatively by Dostal and Mark, formulas have been derived therefore and compared with the experimental results. It seems that only in very special cases one can reduce the whole complicated polymerisation process on these two steps alone, mostly it is necessary to take into account a third elementary effect which interferes with the two steps mentioned, namely, the breaking off of the growth.

*c. Breaking-off processes and sudden finishing of growth.* Our general knowledge of addition reactions between molecules shows that with increasing size of the reacting particles the probabbility of a successful collision, i.e. the addition of a new member in the chain, decreases very rapidly with increased size of the molecules, even if the activation energy remains quite constant. The investigations of Evans, Eyring, Hellmann, Polanyi, Syrkin and Wassermann show that the reactivity falls off very quickly, even when the molecular weight is of medium size (200–400). It is therefore to be expected that the rate of growth should be very dependent on the actual chain length, an influence which was included in the formulæ by Dostal and Mark, with the result that the growth of an individual chain slowly ceases. Comparison with the experiment showed that under certain conditions the polymerisation of styrene seems to be remarkably influenced by this peculiar effect.

On the other hand one must be aware of the fact that a third type of reaction finishes the growth of the chains by annihilating the active groups at their ends. Thus we have to introduce a third reaction constant $k_3$, which summarises all possibilities for the saturation of the reactive spots. They are rather numerous and it is therefore difficult to get a clear idea of the chain breaking.

(1) When the chains are flexible it may arise that the two free valencies on their ends react with one another, producing a ring-molecule of many members. Such molecules are known from the work of Müller, Ruzicka* and others, and even the kinetics of their formation was cleared up to a

* L. Ruzicka : Helv. Chim. Acta ix. 230, 1926.

certain degree by Freundlich and Salomon.* One does not yet know exactly how much this factor influences a given polymerisation reaction, because we have no effective means of finding out how many ring molecules are present in the product of a special polymerisation reaction. In this direction new experimental evidence is wanted ; such a breaking-off reaction would be monomolecular.

(2) A growing system can lose its active ends, if a hydrogen atom wanders along the chain and if a double bond is formed at one end :

$$-CH_2-CHX-\ldots-CH_2-CHX-\ \rightarrow CH_3$$
$$-CHX-\ldots-CH=CHX \quad (c)$$
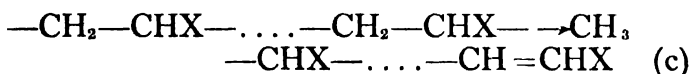
The probability of such an isomeration will be in inverse proportion to the actual length of the chain, and a breaking-off reaction of this type will be of first order, short chains being in greater danger of undergoing this isomerisation than longer ones. Actually in the case of short chains ($n = 2$ or $3$) the final products of this reaction could be isolated in one or two cases, but it is not yet clear which role this type of chain cutting plays in the whole process.

(3) The annihilation of a single active group at one end of the growing, thread-like molecule can also be accomplished during a collision after the equation

$$-CH_2\ldots CHH*-CHX-+M\rightarrow-CH_2\ldots$$
$$CH=CHX+MH* \quad (d)$$
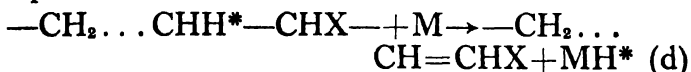
Here the hydrogen atom H* jumps under the influence of the collision with M from the penultimate C-atom to the colliding molecule M and a double bond is formed. Several types of this reaction have been discussed by Dostal, Flory, Mark

* G. Salomon : Trans. Far. Soc. xxxii. 146, 1936.

and Schulz. The probability of such a process is proportional to the concentration of the unsaturated ends and to the concentration of M. The latter partner may be any molecule able to take a hydrogen atom and present in the reacting system, e.g. a normal monomeric particle, another growing chain, a solvent molecule or any impurity of the solution. Even the wall of the vessel could produce the change expressed by the relation (d). Therefore it is not easy to say precisely what will happen in a given case.

An important and interesting question is what becomes of the energy which is set free by the process (d). As a new C-C bond is formed we may estimate the energy excess of (d) to about 70,000 cal. per mole. It may remain in the chain and increase its internal vibrations and rotations, perhaps with the effect that after a certain average life-time also the second end of the chain is stabilised by a similar act. But the energy may also be transferred to M and create a highly active particle. If M is a monomeric molecule it can be transmuted into a new germ, so that one chain is ended by (d) but another is started. There are signs that something like this happens under certain experimental conditions. If M is a solvent molecule the energy may be dissipated by consecutive collisions, but it may also be attributed to one monomeric particle under reproduction of a germ with the aid of the solvent. In recent papers by Dostal, Flory, Mark, Schulz and others these various possibilities are discussed and brought into contact with the experimental facts.

The result is that at the present moment no polymerisation process can yet be resolved quite clearly in all its elementary steps, but there is no

doubt that the three above-mentioned processes

> chain start
> chain growth
> chain ending

always play the outstanding role during a poly-merisation reaction. It would lead much too far to quote all the detailed results which have been obtained up to date, but there is no doubt that the increasing experimental material will enable us to separate more and more neatly the different poss-ible steps and to work out something like a fine structure of polymerisation reactions.

## POLYCONDENSATION REACTIONS

This is the second type of reaction by which high polymers can be produced. Here we will also confine ourselves to reactions which lead to pure chain polymers. In a condensation reaction a new molecule is built under the retirement of a part of the reacting particles ; this part is mostly water. If a two basic acid, e.g. succinic acid, reacts with glycol one gets

$$HOH_2C—CH_2OH + HOOC.CH_2.CH_2.COOH \rightarrow$$
$$HOH_2C.CH_2O.OC.CH_2.CH_2.COOH + H_2O \text{ (e)}$$
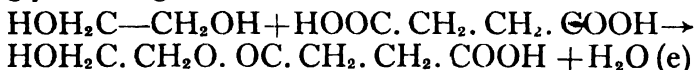
This esterification leads to a molecule which can again react with an alcohol or with an acid, again producing reactive molecules, and growing thus slowly to a long chain with alternative alcohol and acid members ; reactions like this have been investigated recently by Dostal, Flory, Marecek, Raff and others.

They are much easier to study, because the intermediate products are not unsaturated but of the same type as the monomeric substance. It is no chain reaction therefore which we have before us, but a stepwise esterification for which kinetical

formulas can be derived without serious difficulties. The problem of the mutual interaction of three different elementary processes as it has to be solved in the case of polymerisation, is here reduced to the mathematical study of a step reaction with many steps. The production of "germs" is here of the same order of magnitude as the velocity of growth, and therefore the chains which are formed are much shorter than in the case of polymerisation. Therefore fairly good agreement was obtained by comparing the results of the experiments with the above-mentioned formulæ, and the problem of polycondensation reactions does not offer serious difficulties for a quantitative understanding.

This is the reason why our endeavours are more and more concentrated on the study of polymerisation processes, which still offer a great number of unsolved and interesting questions.

[H. M.]

## Friday, November 19, 1937

VISCOUNT FALMOUTH
*Manager and Vice-President, in the Chair*

SIR WILLIAM BRAGG, O.M., K.B.E., D.Sc.,
Pres.R.S., M.R.I.
*Fullerian Professor of Chemistry*

## CLAY

IN a set of Christmas lectures which I gave in 1925–26, I spoke of the changes in old trades due to the impact of new knowledge. A part of the new knowledge consisted in an improved understanding of the materials of each trade. Thus the smith had come to know more about the constitution of his alloys, and the textile worker about the details of his fibres. But the case of the potter was an exception. There was no simple answer to the question "What is clay?" No one could then explain the plasticity of clay nor its power of absorbing water, properties which must depend on the arrangement of its component atoms and molecules. The new X-ray methods had shown that clay was crystalline, and that its structure was neither that of alumina nor that of quartz, although the composition of clay was such that those two substances might have been supposed to be constituents. At that time the X-ray methods had not been so far developed as to be capable of dealing with so difficult a material.

The position has changed. The X-ray methods have been successful in revealing the principal features of the clay structure, though many puzzling details still await explanation. It is now possible to give to those who use clay for their various purposes some explanation of its properties.

The uses are many and the properties curious. The potter finds that his clay, when carefully prepared, can be made to flow easily into his moulds, and that a very slight departure from the customary treatment will ruin his work. The soil chemist finds that clay may vary in its behaviour as an agricultural material through the facility with which it changes its " base " constituents for others: potassium for calcium, for example. He is interested also in its power of absorbing such large quantities of water that its condition may vary from that of a hard solid to that of a sludge. The moulder wants his clay to be plastic while yet it does not wet his fingers. The geologist is concerned with the differences between the many varieties of clay, since he is thereby enabled to picture the conditions when each was formed. Lastly, the crystalline structure itself, and the physical and chemical conditions of the crystals are most important from the scientific point of view.

The recent determinations of the clay structure are based largely upon a series of researches carried out in Manchester by my son, W. L. Bragg*, and his colleagues. Some years ago my son set out to examine the structures of the silicates and with the aid of collaborators succeeded in discovering the principles and the chief details of their construction. The work has been described in the Proceedings of the Royal Society and elsewhere. My son has also spoken of it in this theatre. The next important step was made by Pauling†, who showed that talc, pyrophyllite, the micas and other minerals of the group to which the clays belong, were " layer " crystals, in which sheets of atoms were

---

* See, for example, *The Crystalline State*, W. L. Bragg, pp. 131-136
† *Proc. Nat. Acad. Sci*, Vol xvi, p 123, 1930

piled up one upon another. He measured and described the principal features of their structure. Other workers to whom I must presently refer have advanced further on the lines laid down by Pauling.

Let me remind you that the silicates form a large fraction of the surface materials of the earth. The two elements oxygen and silicon comprise between them about three-quarters of the earth's crust, and therefore it is not surprising that compounds of these elements should also be common. The relative abundance of some of the elements is shown in Table I, which is extracted from a fuller table due to V. M. Goldschmidt.*

## TABLE I

Quantities of elements contained in the earth's crust, in the order of their abundance expressed in grams per ton.

| | | |
|---|---|---|
| Oxygen 494,000 | Titanium 6,300 | Fluorine 270 |
| Silicon 276,000 | Manganese 930 | Chromium 200 |
| Aluminium 88,200 | Phosphorus 786 | Zirconium 190 |
| Iron 51,000 | Sulphur 500 | Copper, nickel |
| Calcium 36,300 | Chlorine 480 | and vanadium |
| Sodium 28,300 | Strontium 420 | each 100 |
| Potassium 25,900 | Barium 390 | Tungsten 69 |
| Magnesium 21,000 | Rubidium 310 | |

All other elements very small, *e.g.* Tin 40, Lead 16, Gold 0.005.

In the silicates the silicon atom is found to be surrounded always by four oxygen atoms, constituting a tetrahedral arrangement with the silicon at the centre. This is the important and persistent feature of the silica structures. The length of the edge of the tetrahedron is always close to 2.55 Angstrom units†.

* *J Chem Soc*, 1937, p 656.

†A centimetre contains 100,000,000 Angstrom units For the sake of brevity the number expressing a quantity in these units is followed only by the letter A , or is left by itself if the meaning is clear.

This association of a silicon atom with four oxygen atoms ranged about it is in accord with the rules that are found to govern the groupings of atoms to form molecules. According to these same rules the clay structures are built up, and as I would like to show how this comes about I will remind you of those sections of the rules which are sufficient for my purpose. My chemist friends will, I hope, forgive this recapitulation on the ground that I must make my present account intelligible to those in the audience who are not familiar with chemical laws.

According to a description which has been and still is extraordinarily useful, though it looks a little crude in the light of modern knowledge, electrons carrying negative charges circulate about the positive nucleus of each atom. The number of electrons is normally sufficient to balance the positive charge, but it can be altered. If it is increased the atom acquires a negative charge, and vice versa. Alterations are generally due to a tendency displayed by these electronic arrangements, according to which certain arrangements are preferred. In particular, there is a much favoured arrangement in which eight electrons placed at the corners of a cube form the exterior shell of the atom. This is notably the case in respect to the atoms which form the clays.

Thus the oxygen atom possesses eight electrons normally, the number being just enough to balance the nuclear charge. Two of these, however, are held firmly in an inner shell, closer to the nucleus. Six only are left to form the outer shell, and the oxygen atom will take over the extra two from other atoms if it has the opportunity, even though in doing so it becomes electrically charged. On the

other hand the metal atoms possess one or more electrons above the number required to complete the cubic arrangement. Thus, the sodium atom has eleven in all : two of these form the inner shell, eight fill the corners of the cube, and one is left over, which, though normally held with its fellows by the elevenfold positive charge on the sodium nucleus, is liable to removal, leaving the atom as a whole with a single positive charge, i.e., a positive charge sufficient to balance the negative charge of one electron.

If one oxygen is brought into association with two sodium atoms, it can gather to itself the two electrons which are held more or less loosely by the sodium atoms. The negative oxygen " ion " now attracts the two positive " ions " and the three form a combination known as a sodium monoxide molecule, having a certain permanency. The elements of common occurrence in the clays are set out in Table II, with the number of electrons that each wants or is ready to part with : –

## TABLE II

| Oxygen − 2 | Silicon + 4 | Sodium + 1 |
| Hydroxyl − 1 | Aluminium + 3 | Potassium + 1 |
| | Calcium + 2 | Hydrogen + 1 |
| | Magnesium + 2 | |

From this table it is clear that a molecule may be formed from one oxygen atom and one calcium atom ; it is the molecule of lime. One oxygen and one magnesium make magnesia. One oxygen, one sodium and one hydrogen make caustic soda. In structural arrangements an oxygen is so often found in combination with one hydrogen that the pair is known as " hydroxyl." The hydrogen

nucleus is buried in the electron sphere of the oxygen, and the hydroxyl is therefore pictured as approximately spherical. The hydroxyl is charged negatively on the whole : like many other molecules it is polar because the opposite charges are not concentric. The combination of an oxygen and two hydrogens forms, of course, the water molecule, which no longer seeks to attach further electrons to itself, nor is willing to part with any.

Imagine now a group consisting of a silicon atom surrounded tetrahedrally by four oxygens. The silicon gives up the four electrons which can be taken from it, and the four oxygens each take one of those electrons, but want one more. Add a hydrogen to each oxygen and the group is satisfied. The oxygens have become hydroxyls. The group now forms the well-known molecule of silicic acid $Si\ (OH)_4$. It is another example of the way that molecules more or less permanent are made. The reasons for the permanency are, in the first place, that the give and take of electrons between the constituent atoms has satisfied the rules concerning electron arrangement, and, in the second place, that the silicon is now charged positively and holds firmly the four negatively charged hydroxyls which surround it. Any atoms that can part with one electron, such as sodium or potassium, would have done as well as the hydrogen.

Let us now develop the silicon-oxygen combination. Imagine a combination of two tetrahedra having one corner, and an oxygen at that corner, in common (Fig. 1). The other six oxygens must be supplied with one electron each, by adding hydrogens or their equivalent. The molecule would then be described as $Si_2O_7H_6$. Carry the combination a stage further. Imagine a chain of tetra
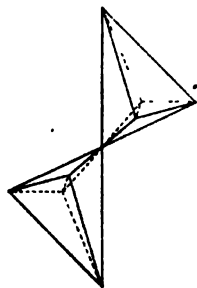
Fig 1. An arrangement of two tetrahedra having a corner in common. They are drawn in full line One is the inverse of the other with respect to the common corner, but this feature of the arrangement is not necessary to the argument. The corners of each tetrahedron may be supposed to be the centres of oxygen atoms which touch each other, and the two tetrahedral cavities so formed may each be occupied by silicon atoms, the centres of which lie at the two intersections of dotted lines.

Fig. 2. A chain of tetrahedra, consecutive links having one corner in common. One face of each tetrahedron lies in a common plane represented as the surface of a rectangular board, and all the tetrahedra are drawn as lying on the same side of that plane, but this particular arrangement is in no way essential

hedra as in Fig. 2, each tetrahedron sharing one oxygen with each of its neighbours. For every silicon there is now one oxygen of the sh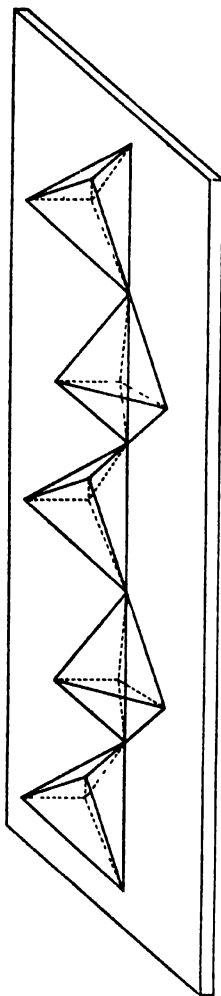ared type, and two that are not shared. These last two need to be supplied with electrons, which may be done by adding to each a hydrogen, sodium or other monovalent atom ; let it be sodium. The result is a chain of indefinite length, consisting of links each repeating the combination $Na_2SiO_3$. The ends of the chain, of course, require finishing off by the addition of a monovalent atom to each of them.

This combination is known as a sodium silicate. It has a strong affinity for water, as clay has ; and probably the structural feature to which this is due is the same in the two cases. Thus surrounded with water molecules, it makes a viscous medium known as water glass ; the viscosity is due to the entanglement of the long chains.

Having now repeated the tetrahedra in one direction to make chains, let us next expand the chains sideways so as to make sheets. Such an arrangement is shown in Fig. 3.* The complete circles in full line represent hydroxyls, under each of which lies a silicon ; this again rests on three oxygens. Silicons and oxygens are shown in dotted or partly dotted line. The rules of composition are obeyed because each silicon contributes four electrons to the three oxygens and one hydroxyl which it touches, and each oxygen draws an electron from each of the silicons which it touches. Each silicon has one hydroxyl to itself and has half shares in three oxygens, so that the composition is $SiH (O + 3/2O)$ or more simply $Si_2O_5H_2$. A photograph of this arrangement is shown in Plate IA. Of course, the sheet has a boundary at which something special

* See also *The Crystalline State*, W L. Bragg, p. 135.

must be done, but when the sheet is of very great
extent and contains many thousands of atoms, the
fringe is relatively unimportant and does not enter
into our present considerations.

Such a sheet has been termed a hydrated silica
sheet ; in what follows we may refer to it briefly as
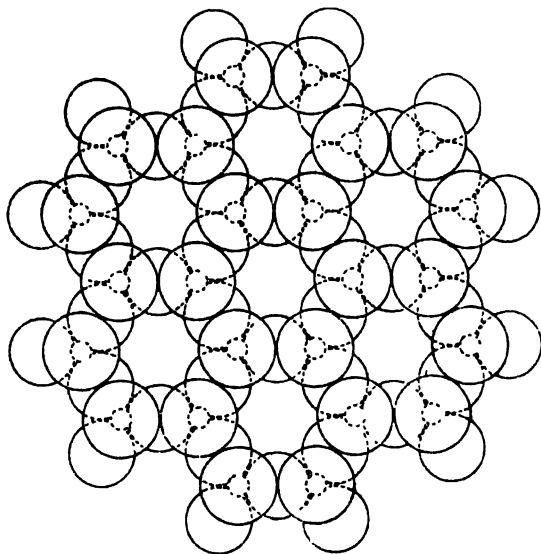


FIG. 3. In this figure the completed circles in full line represent
hydroxyls, each of which rests on three oxygen atoms, shown
smaller (ratio 2 : √3) and partly hidden  The four enclose a
cavity within which lies a silicon atom. Each tetrahedron now
has three oxygen atoms in common with neighbouring
tetrahedra. The tetrahedra all lie on the same side of a plane
in which lies one face of each tetrahedron. See also Plate I.

a " silica sheet." It is not found in individual
existence, but it is of frequent occurrence in com-
binations as will be seen presently.

From the two-dimensional sheet we might go
on to the three-dimensional solid. Thus in quartz,
tridymite and cristobalite, every oxygen belongs to

two tetrahedra, hydrogens no longer enter into the composition. The structures thus laced together are hard and strong. The three-dimensional structures of this kind are outside our present consideration, however. We proceed no further than the
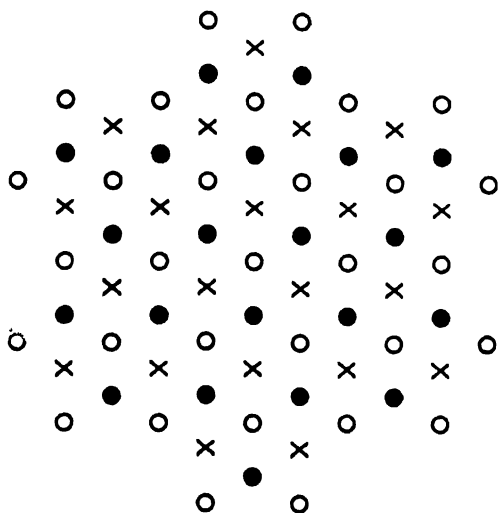


Fig. 4. The black dots represent the centres of a sheet of hydroxyls lying in a plane. The white circles represent a second sheet lying upon the first. In both sheets, each hydroxyl is in contact with six neighbours in its own sheet, and three in the other sheet. The crosses represent the centres of magnesium atoms, each of which is enclosed in the octahedral cavity formed by six hydroxyls, three from each sheet. See also Fig. 6.

sheets, which form an important constituent of the clays.

We must now turn our attention to combinations of oxygen with aluminium and magnesium. Let the dots in Fig. 4 represent the centres of hydroxyls lying in a plane, the hydroxyls being supposed to be spheres in contact, each touching six neighbours. Let magnesium atoms be placed in the positions

represented by the small crosses, each settled in the triangular-shaped cavity between three neighbouring hydroxyls. It will be observed that there are as many crosses as dots, since a cross lies just below every dot. Next place a second layer of hydroxyls over the first, their centres being—in plan—in the positions shown by the small circles.
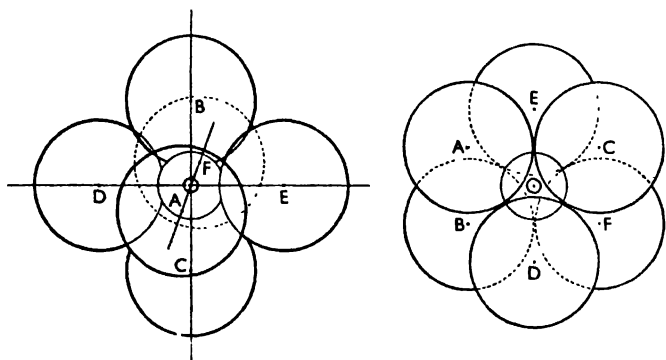


FIG. 5. Two views of a set of six spheres lying at the corners of an octahedron, each sphere touching four others. The left-hand figure shows best the octahedral arrangement, the right-hand the disposition of any set of three in contact to the other set of three. A seventh sphere is shown in the octahedral cavity. The six large spheres might be hydroxyls and the enclosed sphere a magnesium. Actual trial is helpful.

Each magnesium is thus enclosed in the cavity formed by the six hydroxyls which touch it, three from below and three from above. The six hydroxyls lie at the corners of a regular octahedron. This may not be obvious at first, but is readily appreciated if consideration is given to the diagrams of Fig. 5, which show the octahedron from two points of view. The radius of the magnesium must, of course, be much smaller than that of the hydroxyl in order to fit into the cavity ; the ratio of the two radii must be less than 0.41.

In this double sheet of hydroxyls enclosing a single sheet of magnesium atoms, there are two hydroxyls to every magnesium. Since each hydroxyl demands one electron and each magnesium has two to spare, the combination rules are satisfied. A pile of these compound layers forms the crystalline material " brucite " $Mg(OH)_2$. The crystals are hexagonal, as might be expected from the arrangement of the component atoms.

If aluminium atoms are substituted for magnesium atoms in this arrangement, the new formation is called " gibbsite," or " hydrargillite." The aluminium atom has three electrons to give as compared with the magnesium's two, so that the number of atoms of the former kind is only two-thirds of the number of magnesium atoms in brucite. The arrangement is shown in Fig. 6.

If the arrangement of metal atoms in Fig. 6 is compared with that of Fig. 4—crosses in Fig. 4 and black spots in Fig. 6—the horizontal lines being taken as the objects of comparison, it is clear that every third atom of the brucite corresponds to an empty space in the gibbsite.

The clay known as " halloysite " is composed of a pile of gibbsite and silica sheets in alternation. The adjustment of each layer to its neighbours must be made according to some plan ; we must now give some consideration to the details of that planning.

In the hydrated silica sheet the atoms are tied firmly together by the electrical forces which arise from the transference of electrons from atom to atom. The same sort of binding holds the gibbsite layer together. But there has been no transference of electrons from the silica sheets to the gibbsite sheets, or vice versa. Each sheet is neutral as a

whole. Consequently there are no binding forces to draw the two kinds of layer together of the kind that draw together the atoms within the layers. How then do the two layers attract each other ?

The silicate and gibbsite layers are neutral, each containing an equal number of positive and negative charges. Nevertheless, lines of force emerge
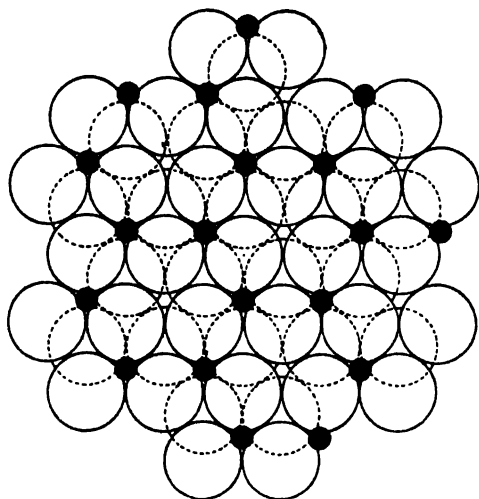


FIG. 6. This is the same arrangement of hydroxyls as in brucite, but the aluminium atoms are fewer than the magnesium in the ratio of 2 : 3. This crystal is not quite hexagonal; it has been measured by Miss Megaw, Zeit. f Krist. Vol. liiivii. p. 185.

from their surfaces to return elsewhere. If a geometrical surface be drawn close to a layer, but so that all the charges lie on one side of it, as many lines of force cross the surface one way as the other. The surface is equivalent to a patchwork of electrical charges. We may roughly represent such a state of things by a packet of magnets stuck into a board, as in the figure (Fig. 7) ; half the magnets point one way and half the other.

If a second set of magnets B is brought up to the first, which we denote by A, the two fields of force intermingle. If the approach is made in certain ways there is attraction, if in others which alternate with the first, there is repulsion. If A is slung by
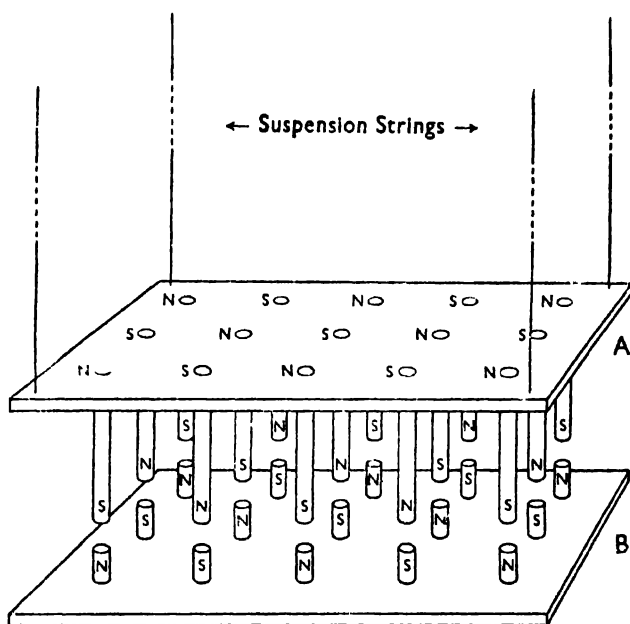


FIG. 7. Magnets are set into two boards A and B as shown in the figure. The board B is fixed, A is slung by long cords. The magnets alternate in polarity. The regularity of the figure is not however essential, nor need the settings in the boards be the same. Whatever the settings may be, forces will act between the two boards, and A will take up one or other of certain positions of equilibrium.

cords so that it can only move from side to side, it takes up one or other of certain possible positions of equilibrium to which it returns if it is displaced sideways. If an experiment could be arranged in which A and B were silica sheets or gibbsite layers,

or any such sheets or layers, the two being alike or different, the same effect would be found. In all cases, however, there is a limitation of great importance which we must now consider.

In each crystal sheet there is a two-dimensional unit of pattern. Thus, in the silica sheet of Fig. 8 A it is shown by the rectangle ABCD, in the other (Fig. 8 B) by A'B'C'D'. In the former case the whole sheet can be divided into similar rectangles, each having the same content and arrangement of contents as ABCD. A similar division can be made in the latter case. Now it happens that ABCD and A'B'C'D' are very nearly equal to each other in all respects : for our purpose we may assume that they are exactly equal. Suppose now that in our imaginary experiment the mutual arrangement of the two layers, silica and gibbsite, is such that the portions within areas ABCD of the one and A'B'C'D' of the other, the one area overlying the other, are in stable equilibrium ; that is to say, if one were displaced sideways with respect to the other it would tend to return. Then the same holds rectangle by rectangle all over the sheets; the effect is cumulative. The two layers as wholes are in stable equilibrium.

But if there were any difference in the dimensions of the two units of pattern, then in some parts of the surface of separation there would be a tendency to reverse any displacement that took place, and in other parts to increase it. Forces would cut each other out, and on the whole the two layers would have no influence on each other.

In the model (Fig. 7) the upper sheet A is restrained from direct approach to B : it can only move sideways. Thus the analogy with the actual case is incomplete. But we can easily imagine what must
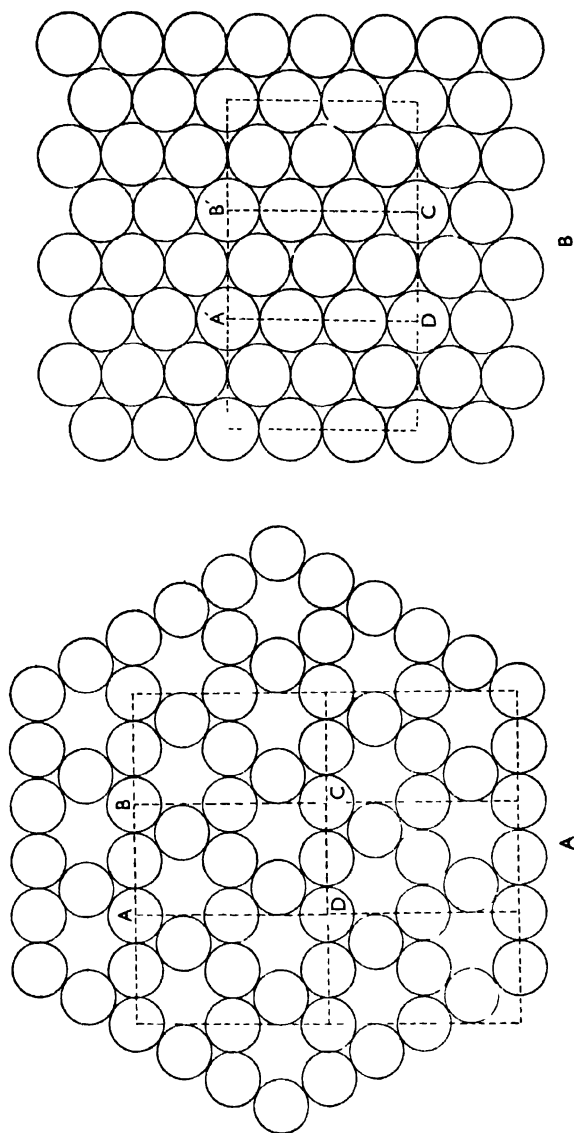
FIG. 8. These two figures show (A) the pattern of the oxygen sheet in the silica layer and (B) the pattern of the hydroxyls in the gibbsite layer. The rectangular unit of pattern is shown in dotted lines and has the same dimensions in the two cases. In each case it encloses six atoms, in wholes or in parts

happen in the case of the two molecular layers when their movement is perfectly free. They will be drawn together and will take up a position—it may well be one of several possible positions—in which the electrical fields best fit together ; or in other words, one of several possible positions of minimum potential energy.

The whole formation, therefore, depends on the exact equality of the two rectangular units of pattern. The two crystalline layers must fit perfectly in combination, even if they differ slightly in separate existence. It is not difficult to imagine that such a difference could be overcome. Perhaps an enforced fit would more readily take place if one of the two layers were laid down on the other, atom by atom, or if the two grew together. Evidence that this adjustment actually happens will presently appear.

The lengths of the sides of the rectangle are in each case 5.1 and 8.8 very nearly. The ratio of these two is 1 : $\sqrt{3}$ as it ought to be, in accordance with the diagrams of Fig. 8. The shorter side is twice the diameter of the oxygen atom, the longer is three times the size of the hydroxyl, the diameters of which are therefore 2.55 and 2.93. In the case of gibbsite these values are determined accurately by the X-ray methods.* In the case of the hydrated silica layer they are not observed directly but are inferred, partly from the known dimensions of the quartz structures, partly from the X-ray determinations of the structures of the clays themselves, of mica, talc and other crystals in which the silica layer is found.

The structure of the halloysite crystal has been determined by Mehmel.† The dimensions of the

* Megaw, *Zeit. f. Krist.*, Vol. lxxxvii, p. 185, 1934.

† *Zeit. f. Krist* , Vol. xc , p. 35, 1935.

unit cell are $a = 5.20$, $b = 8.92$, $c = 10.25$, $\beta = 100°$. The values of $a$ and $b$ correspond closely to those of the separate gibbsite sheet. The value of $c$ is of the order to be expected if it measures the thickness of four oxygen or hydroxyl sheets piled one upon the other, two such sheets belonging to the silica and two to the gibbsite. The X-ray measurements show that the thickness of the latter two is 4.86 : and the thickness of the former must be nearly the same. The complete list of the contents of the unit cell can be set out as in Table III. The lowest sheet of oxygens in the unit cell contains six oxygens within the rectangle $a \times b$, as may be seen from Fig. 8A. In the next sheet four silicons lie within the cell and then come four hydroxyls. Above these lie in order six hydroxyls, four aluminiums and finally six hydroxyls. The contents of the cell may be described by the formula $Al_4Si_4O_{22}H_{16}$. Alternatively the contents may be described as $2Al_2O_3 \cdot 4SiO_2 \cdot 8H_2O$ ; but it is to be remembered that, though this latter expression conveniently describes the contents in terms of known molecular forms, those forms lose their identity in the crystalline structure.

When halloysite is heated slightly, its two middle sheets of hydroxyls collapse into one, consisting of a sheet of four oxygens and two hydroxyls. Four oxygens and eight hydrogens are expelled, so that we may describe the process as an expulsion of water, even though there were no distinct molecules of water in the halloysite. In the new form of crystal, known as meta-halloysite, the oxygens in the middle sheet lie over the silicons, and the hydroxyls in the same sheet lie over the cavities in the bottom layers of silicon and oxygen. It is easily seen that the proper electronic relations are fulfilled.

## TABLE III

Arrangement of the layers of atoms in clays and allied substances. The unit cell of the crystal consists of sheets of atoms packed upon one another: each sheet being enclosed within a rectangle about $5.2 \times 8.8$A. The height of the cell depends mainly on the number of oxygen or hydroxyl sheets. The numbers attached to the sheets in the table give the number of atoms or hydroxyls in each sheet. An approximate value is given for the thickness of each stack, *i.e.*, from the beginning of one stack to the beginning of the next.

| HYDRATED SILICA | BRUCITE | GIBBSITE |
|---|---|---|
| 4 OH<br>4 Si<br>6 O | 6 OH<br>6 Mg<br>6 OH<br>Thickness = 4.73 | 6 OH<br>4 Al<br>6 OH<br>Thickness = 4.86 |

| HALLOYSITE | KAOLINITE |
|---|---|
| 6 OH<br>4 Al<br>6 OH<br>4 OH<br>4 Si<br>6 O<br>Thickness = 10.25 | 6 OH<br>4 Al<br>4 O+2 OH<br>4 Si<br>6 O<br>Thickness = 7.2 |

| PYROPHYLLITE | MONTMORILLONITE | MICA (Muscovite) |
|---|---|---|
| 6 O<br>4 Si<br>4 O+2 OH<br>4 Al<br>4 O+2 OH<br>4 Si<br>6 O<br><br>Thickness = 9.4 | 6 O<br>4 Si<br>4 O+2 OH<br>4 Al<br>4 O+2 OH<br>4 Si<br>6 O<br>Water layer or layers<br>Thickness 9.2—21.4 | 6 O<br>3 Si+Al<br>4 O+2 OH<br>4 Al<br>4 O+2 OH<br>3 Si+Al<br>6 O<br>2 K<br>Thickness = 10.0 |

The interest of the meta-halloysite structure lies mainly in the fact that it differs very little indeed from that of the well-known kaolinite or china clay. The X-ray methods show that the composition and arrangement of the layers is the same. But there are slight differences between the X-ray photographs of the two substances which may possibly be due to slight sideways shifts of the two layers relatively to one another, or to slight displacements of the atoms in the layers.

Nacrite and dickite are two other forms of clay very closely allied to kaolinite, differing therefrom in the same way—in all probability—as meta-halloysite. A set of photographs of some varieties of clay is shown in Plate I B. The many points of agreement between them are obvious. So also are slight variations principally in the intensities of the lines ; these small differences still require explanation.

Now we come to the second and only other group of clays. Chemical analysis shows that in general the proportion of silicon to aluminium is much higher in this group than in the first. There is, however, much variation in composition, and in fact the differentiation between various types and individual members of the group has been a difficult problem. The problem has now been solved, so far at least as its main points are concerned.

The typical feature of the group is the fact that there are two silica layers to one layer of the other kind, brucite or gibbsite, or modifications thereof : there are two tetrahedral layers of oxygen or hydroxyl to one octahedral layer of the same components. It is generally accepted that the latter layer lies between the two former,* so that the

*Pauling *loc. cit.* See also Gruner, *Zeit f. Krist.*, Vol. lxxxviii, p. 412, 1934.
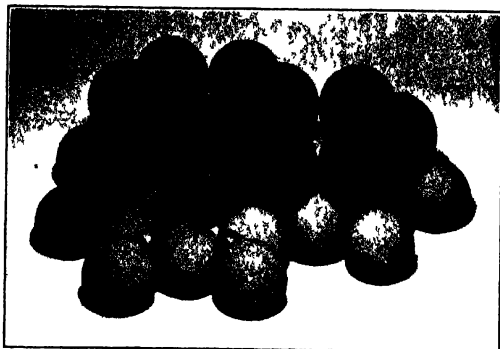
**A** A photograph showing the arrangement of the atoms in a hydrated silica sheet The oxygen atoms that form the lowest sheet are arranged as in Fig 8A The hydroxyls, shown darker, rest on the oxygens in the arrangement of Fig 8B Silicons occupy the tetrahedral cavities two cavities are opened up to show these
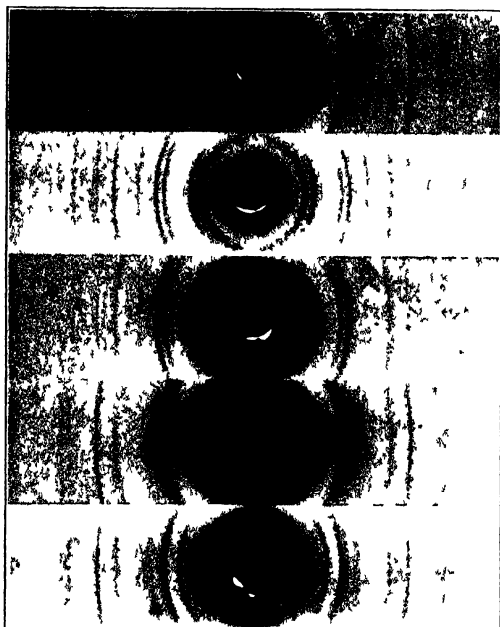


NACRITE

DICKITE

KAOLINITE

HALLOYSITE

MONTMORILLONITE

**B** A set of X-ray photographs of a few types of clay These are due to Dr Nagelschmidt

PLATE I

arrangement is symmetrical and non-polar. Evidence on this point has been accumulating, and is now strong. The arrangement of the layers is shown in Table III. Pyrophyllite may be taken as the type of the group, which contains montmorillonite, beidellite, bentonite, fuller's earth and other varieties. The micas, talc, clintonites and others of this kind, though their properties differ markedly from those of the clays, are extreme varieties of this group.

Since the individual layers are the same as in the first group, and the difference between the two groups consists only in the order and relative proportions in which they are piled on one another, it is to be expected that the X-ray photographs will be similar in many respects. This may be observed in the photographs of Plate I B. All the lines due to sets of planes which contain the c-axis of the crystal, planes which therefore depend only on the form and dimensions of the unit of pattern in the layer, must be the same throughout. The principal difference is in the spacing along the c-axis, which depends directly on the number of layers ; it is nearly proportional to the number of sheets of oxygens or hydroxyls. In fact, the length of the c-axis is used as a means of counting the layers.

The differences between the members of the pyrophyllite group are shown in variations of their metal contents. We shall find it convenient to study them in connection with the important phenomenon of base exchange. Meanwhile, one variation is of immediate interest. When the gibbsite is replaced by brucite, the new substance is talc. It is natural to ask why there is not the same parallelism in the first group. Why is there no variation of kaolinite in which the same substitution

has been made ? Pauling has suggested an in-
teresting explanation. The unit rectangle of the
magnesium crystal brucite is appreciably different
from that of the silica layer. There is an approxi-
mate 8% increase in the dimensions as compared
with the same dimensions in the aluminium crystal
gibbsite. Where silica layers lie on both sides of
the brucite, as in talc, the adjustment may be made
forcibly ; but this cannot be done in the kaolinite
type, where there is only one layer of each kind.
The argument supports the hypothesis that the
arrangement of tetrahedral and octahedral layers
in pyrophyllite is a symmetrical one.

These then are the clay structures. We have
now to consider whether any of our new knowledge
furnishes explanations of well-known properties of
clay. It is true that the details of the structures
have not yet been worked out in full, and our
explanations cannot therefore be complete. But the
general features are so clear that it becomes
interesting to enquire how far we can go.

Let us consider first a well-known property that
clay particles show in common with many other
colloids, namely, the electric charge of the particles.
In the case of clay the charge is almost always
negative. The fact is easily made the subject
of a lecture demonstration. Two electrodes of
platinum sheet are dipped into a narrow cell placed
in front of the lantern and show as fine lines upon
the screen. The cell is filled with a suspension of
finely divided kaolin and a suitable electric poten-
tial is imposed. In a minute or two it becomes
obvious that the space about the negative electrode
is being cleared of clay particles. (Plate II).

The clay particle is normally uncharged. But as
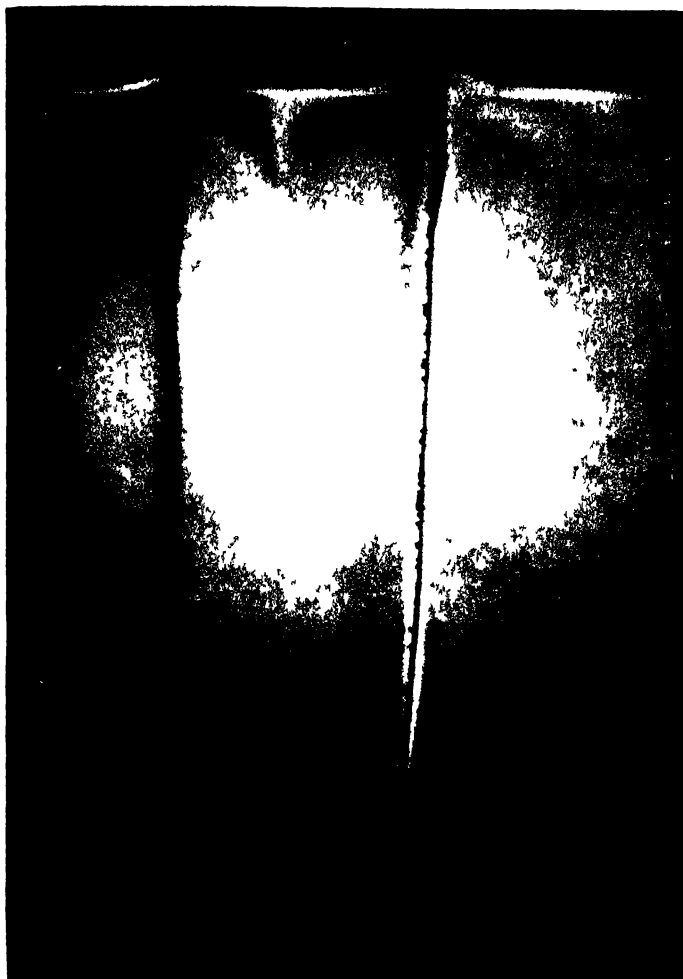I have said already, its surface, or any sheet drawn

PLATE II A photograph showing the appearance upon the screen when a projection is made of a cell containing a suspension of fine clay particles, and containing also two electrodes There is a clear space about the negative electrode

about it like an envelope, is crossed and recrossed by lines of force. Lines emerge, curve round and cross again, the emergence spots acting as a positive charge and vice versa. The pattern is altered by the approach of other fields, the result being a combination of fields. The atoms of the crystal, to which the crystal field is due, may alter their dispositions, but they form part of a locked system, so that such movements are small—if they were so large as to displace the atoms altogether there would be " chemical action "—and have only a secondary influence. Just so when one magnet approaches another the two sets of lines form a new figure ; but where the magnets are separated again the magnets reform their individual fields and are as they were before the approach.

Such a crystal surface has a variety of local attractions and repulsions for charges that approach it. If there are ions, i.e., charged atoms or molecules, in a surrounding fluid, positive ions will be attracted to the negative spots on the surface, and vice versa. If the two kinds of ions, positive and negative, are attracted and held in equal numbers, the surface remains unchanged. But such equality is not to be expected, because the positives and negatives differ in the forms of their fields and in their mobilities. One is sure to fit better into the surface field and one will have more chances of doing so than the other. The oxygen surface of the clay particle apparently retains the negative hydroxyls more firmly than the equivalent positive ions, and so becomes negatively charged. The positive ions lie loosely in the surrounding liquid. The electrical field of the hydroxyl must be very different from that of a positive ion of the single-atom type, because it

is dipolar, and for that reason may fit on to the oxygen surface of the silica sheet with especial aptitude.

On this view the charge of the colloid particle is mainly determined by structural conditions. It might be either positive or negative, the sign in any given case depending partly on the form of the surface field, and partly on the properties of the ions that in Brownian movement are continually washed up against the crystal.

A clay particle thus loaded with negatively charged hydroxyls repels other similar particles, so that their suspension·in the surrounding liquid is permanent. We know, however, that the repulsion can be removed by supplying ions to the liquid. Thus, a little hydrochloric acid added to the potter's slip coagulates it at once. The negative charges of the ions that lie on the clay surfaces are balanced by the positive hydrogens that settle on them, and the particles no longer repel one another. A relatively small quantity of acid is required because the number of atoms to be deposited is small in comparison with the numbers within the particle itself ; just as the weight of the paint required to cover a house is very small in comparison with the weight of the house.

One of the most remarkable properties of clay is the ease with which water can be added to it or taken away from it, the change taking place within the clay itself. Water can, of course, be added to or removed from a suspension of clay particles, but the addition in question is structural. The clay known as montmorillonite shows this property well, and has consequently been the subject of special examination by Nagelschmidt* and by

* *Zeit. f. Krist* , Vol. xciii , p. 481, 1936

Hofmann and others. Pyrophyllite, as described in Table III, contains four atoms of hydrogen in each unit cell ; they form part of the main structure and cannot be removed without destroying it. The cell may be said to contain two molecules of water, but it must be again observed that the four hydrogens and the corresponding two oxygens are not linked up as they would be in separate water molecules.

The pyrophyllite structure is the same as that of montmorillonite, from which all water has been driven off except of course the hydrogen and oxygen which would make up water molecules if the structure were broken up. The length of $c$-axis is then 9.4 Ångstroms, the value which, as we have already seen, is proper to a pile of four oxygen or hydroxyl sheets piled upon one another and cemented by atoms of silicon and aluminium.

When water is added to this structure, the $c$-axis is increased in length. Nagelschmidt's measurements made in 1936 have lately been revised and extended by Bradley, Grim and Clark,* who observe that there is a well-defined set of hydrates, in which the water contents of the cell increase from two molecules to 8, 14, 20 and 26 molecules successively and that the $c$-axis grows correspondingly from 9.6 to 12.4, 15.4, 18.4 and 21.4. This means that sheets of water molecules are added to the pile in each crystal cell. The average increase agrees with what might be expected on that hypothesis, because the added volume of one sheet is $3 \times 5.2 \times 8.9 - 138.5$, and the added weight is $6 \times 18 \times 1.66$ Angstrom units of mass $= 179.5$. The density of the added sheet of water is therefore $179.5/138.5 - 1.3$ nearly. The size of the water mole-

cule is known from several crystal measurements ; and it can be calculated that such molecules in close packing would give to water a density more than half as great again as the normal value, unity. Water has presumably an open structure. According to Bernal and Fowler,* it tends to assume a quartz-like character ; groups of molecules arrange themselves temporarily and then break away again. The density of the sheet agrees therefore with the conception of an arrangement of water molecules more closely packed than in water itself, but not so closely packed as might be.

If water molecules are thus attached to the clay particle, they cease to belong entirely to the liquid, even where the oxygen sheet on which they are deposited forms the exterior surface of the crystal. The clay has adsorbed them or absorbed them, the former term being the more appropriate when reference is made to the external surface of the clay particle, and the latter when reference is made to the spaces between the oxygen sheets of the silica layers. The looseness of their association gives easy movement to the clay particles, and an easy slip over the surfaces of the layers. But the clay is not wet unless the amount of water exceeds a certain proportion. It is plastic without being sloppy. The comparative looseness of the holding will also allow the ready removal of water by heat or evaporation. The amount of hydration will be in equilibrium with a certain moisture content of the atmosphere.

Thus, the ability of clay to absorb and release a relatively large quantity of water is due to its peculiar structure. Sheets of water molecules can be inserted successively into the layer-like arrangement. They fit into the crystalline structure,

* *J Chem Phys* , Vol 1 , p. 515, 1933

causing an increase in the length of the c-axis, but no change in the lengths of the other axes. This peculiarity is also the cause of plasticity.

We now have to consider one more remarkable property of some clay structures, that of " base exchange." It is of great importance to the agriculturist. The property is possessed by certain other crystalline materials, for example, the zeolites. These crystals contain a proportion of metallic ions, sodium, potassium, calcium and others, the presence of which is due to a special feature of their design, which we must proceed to consider. As these atoms are positively charged, they may be considered to be " bases " linked up to the negative remainders of the crystal. They can in certain cases be exchanged with each other and hence the term by which the process is described.

The zeolite crystals are well formed and easily handled, so that their structures, though complicated, can be examined by the X-ray methods. The structure of analcite, a member of this group, has been worked out in some detail by W. H. Taylor*, and the process by which base exchange is effected is made clear. The clay structures are not so easily measured, but the process is certainly on the same lines.

The special feature in the design of these crystals on which the base exchange is founded is the occasional replacement of the silicon atoms in the oxygen tetrahedra by aluminium atoms ; other replacements may occur but are of less importance. Now, an aluminium atom can supply only three electrons, whereas the silicon which it replaces has been supplying four. Consequently, whenever such a substitution occurs, another electron must

* *Zeit f Krist*, Vol lxxiv , p 1, 1930

be provided from somewhere. It may be supplied by an extra atom of sodium or potassium or calcium. Such an extra atom is not essential to the main structure of linked tetrahedra. Its removal would leave the structure negatively charged, but would not disturb the structural plan seriously. In certain cases it is comparatively easy to change one kind of these balancing atoms for another.

Not all structures so possessing balancing ions exhibit the process of base exchange. It would seem to be necessary also that there should be easy roads by which the ions may go in and out. In the case of analcite, Taylor shows that channels run through the crystal, along which the balancing ions may be supposed to travel. A certain number of water molecules are found in these channels so that, as in the case of the clays but to a smaller degree, the substance can be deprived of water and can recover it.

In the class of silicates to which the clays belong the same substitution of aluminium for silicon is found to take place. Compensation is effected by the addition of sodium, potassium or other atoms which can give up electrons. In the mica structure, for example, potassium is the compensating element, as shown in Table III. In this case it would appear that the potassium is too strongly bound up with the rest of the structure to be easily displaced. The potassium atoms weld the layers together and make them into a firm crystal. They are not strong enough, however, to prevent the ready cleavage so characteristic of mica. When calcium is substituted for potassium, as in the " clintonites," the bonds are stronger and cleavage is no longer possible.

In the montmorillonites, however, the base exchange is easily effected. We may suppose that the

semi-liquid sheets of water which are found in it allow the compensating atoms to possess some freedom of movement. They perform a duty analogous to that of the channels of analcite. It is in the montmorillonite group that base exchange is so common and has such important consequences.

To sum up, it would appear that the structures which are found as the result of X-ray measurements give reasonable explanations of the clay properties. Many details remain to be filled in, principally because it is difficult to obtain pure material to work on, and because the material, though it may be uniform in composition, is complicated by the substitutions which are nearly always present. More complete knowledge is sure to come quickly from the many laboratories in which the clay problem is under consideration ; the general character of the solution is already clear.

This Discourse has been illustrated by experiments showing the electrical charge upon particles of kaolin (Plate II), the behaviour of the particles as crystals in polarized light and the orientation of the crystals in an electrical field. For the skilful preparation of these demonstrations I am indebted to Mr. W. J. Green and his assistant Mr. Bridger. Messrs. Wedgwood provided samples of various clays, including the clay slip with which to show the coagulation on the addition of acid. Dr. Hatschek lent specimens of clay showing thixotropy exceptionally well. [ W. H. B.]

# Friday, November 26, 1937

CLIFFORD C. PATERSON, O.B.E., D.Sc., F.Inst.P.
*Manager and Vice-President, in the Chair*

H. DINGLE, D.Sc., A.R.C.S.
*Assistant Professor of Astrophysics, Imperial College of Science
and Technology*

# SCIENCE AND THE UNOBSERVABLE

A NEW phenomenon has appeared in modern
physics : namely, an attempt to apply with rigour
the principle that only that which is observable is
significant. This is not intended to be a precise
statement of the principle. It is at least vague, and
perhaps inaccurate, but on that very account it is
the most suitable statement with which to begin
our discussion. For the principle itself has not yet
been clearly isolated from its applications. Uncon-
sciously, progressing more by instinct than by
sight, physicists have allowed it to direct their
thoughts, but they have not succeeded in giving it
clear expression. Consequently it is manifested
here in one form and there in another, appearing to
some as an outstanding example of scientific arro-
gance and even absurdity, while to others it has an
authority which raises it above commonsense and
reason alike. The resulting controversy, as may be
imagined, has been not without dust and heat,
some of which, I fear, must be introduced into the
placid atmosphere of this Institution. Let us agree
at the beginning that our own considerations shall
be at least cool. I will try to mitigate the dustiness,
and we will make an unimpassioned attempt to
understand the meaning of the principle and to

reach a viewpoint from which its validity may be justly appraised.

Let me say at once that the principle is not new. I have no doubt that the Greeks had some words for it. It is certainly to be found disturbing the thoughts of Galileo, Newton, Locke, Hume, Kant, Huxley, Mach and many others ; Huxley, for example, speaks of

"that exact verbal expression of as much as we know of the fact, and no more, which constitutes a perfect scientific theory."*

But on very few occasions in the past has it been more than a pious belief, a doctrine which it has been proper —or only slightly improper—to hold, but which not only has not been allowed to influence the actual prosecution of scientific or philosophical speculation, but has actually been violated therein. It is Einstein who is responsible for the importance which it has to-day—not because he has stated it more clearly than others, but because he has deliberately brought it into practice and thereby achieved a striking scientific success. Let us hear him speaking of the crux of his theory—the idea of the simultaneity of events occurring at different places.

" The concept " (of simultaneity), he says, " does not exist for the physicist until he has the possibility of discovering whether or not it is fulfilled in an actual case. We thus require a definition of simultaneity such that this definition supplies us with the method by means of which, in the present case, he can decide by experiment whether or not both the lightning strokes occurred simultaneously. As long as this requirement is not satisfied, I allow myself to be deceived as a physicist (and of course the same applies if I am not a physicist), when I imagine that I am able to attach a meaning to the statement of simultaneity. (I would ask the reader not to proceed farther until he is fully convinced on this point)."†

* *Hume* (" English Men of Letters "), p. 55, 1879.

† *The Theory of Relativity* (English translation) p. 22, 1920.

I think that if this statement had been made at any time in the history of modern science, but had not been applied to the foundations of physical conceptions, it would have aroused no comment. Einstein would have been regarded as simply saying that if our definition of simultaneity is such that it is essentially impossible to determine whether events are simultaneous or not, then it must be wrong; and everyone would have agreed and turned to more important matters. But he went further. He gave a definition which both accorded with scientific practice and satisfied the requirement he had stated, and then went on to show that certain modifications of our generally accepted notions of space and time must necessarily follow. This put the fat into the fire, and instead of being ignored for saying the obvious, he was railed at for saying the absurd. Philosophers in particular were aroused, and when they realised that a mere empiricist was venturing to challenge their fundamental principles in tones which resounded throughout the world, they proceeded to administer the rebuke which the occasion demanded. Here, for example, is a comment by M. Jacques Maritain, the well-known French Catholic philosopher, on the passage I have quoted (in which, it will be remembered, Einstein asks the reader who is not convinced of his principle to read no farther) :

" Let us, then, obey our author and read no farther, for this little parenthesis, ' the same applies if I am not a physicist,' is of direct concern to us who have not the honour to be physicists, and it presumes to introduce us into the most fallacious metaphysics. . . . It is a fault so obvious to the eyes of a philosopher to confuse the *meaning* of a concept . . . with the *use* which may be made of the concept in this or that field of study, and more particularly to confuse a thing . . . with the measure which we take of it by our

senses and our instruments, that we hesitate to impute such a mistake to anyone, whoever he may be. Everything goes to show, however, that Einstein has made this mistake."*

(We may remark in passing that in speaking of *measures*, Maritain has somewhat confused the issue ; measurement is not mentioned by Einstein. The point M. Maritain really wishes to make is that the nature of the thing we observe, whether we measure it or not, is independent of our devices for observing it and should therefore be independently defined.)

It is not surprising that a physicist and a philosopher should take opposite views of this question, but the matter cannot be disposed of on simple psychological grounds. Here are the diametrically opposed views of two men of science :

" The general point of view " (of relativity), writes Prof. C. G. Darwin, a mathematical physicist who needs no introduction to this audience, " of questioning the reality of anything unobservable is one of the greatest revolutions in scientific thought that has ever occurred. . . . The great idea which Einstein contributed to scientific philosophy was the principle that if a thing is essentially unobservable then it is not a real thing and our theories must not include it."†

This remark has caught the eye of Mr. Albert Eagle, now a mathematician but in his better days an experimental physicist, whose name is known wherever spectroscopy is practised.

" To me," says Mr. Eagle, " this ' great idea ' is the most savage example of the application of what is known as the principle of *Occam's Razor* of which I have heard. . . . Einstein's ' great idea ' requires us to surrender our com-
·monsense for the sake of an arbitrary dictum of his which he and his followers have raised to a fetish. It is prepos-
terous, and to my way of thinking is so inherently idiotic that I cannot understand anyone wasting his breath in giving

* *Réflexions sur l'Intelligence et sur sa Vie Propre*, pp. 205–6, 1926.
† *The New Conceptions of Matter*, pp. 23, 81, 1931.

utterance to such a view. If I had first heard this opinion uttered at a scientific meeting by some scientific nonentity I should have longed to have got up and said that in my opinion such a view was simply a bit of perverse imbecility."*

And finally, here is the attitude of another philosopher, Rudolf Carnap—perhaps the leading exponent of the most active of modern schools of philosophy, the so-called "logical positivism." Carnap, the philosopher, not only accepts the principle ; he uses it to reduce to nonsense those branches of his own subject known as metaphysics and ethics.

"I will call *metaphysical*," he writes, "all those propositions which claim to present knowledge about something which is over and above all experience. . . . Metaphysicians . . . are compelled to cut all connection between their propositions and experience ; and precisely by this procedure they deprive them of any sense."†

In these days, when politeness in controversy, so desirable and proper if directed to the disputants, is so often improperly extended to their ideas, so that few dare speak disrespectfully even of the equator, there is something refreshing in the forthright character of these remarks. And not only are they refreshing ; they are highly significant also. When men who are neither fools nor liars agree that a certain idea is either the greatest discovery of a generation or the silliest nonsense imaginable, but cannot agree as to which it is, it is clear that there is more than a difference of opinion ; there must be some defect of understanding also. Intelligent men do not thus differ about that which they comprehend equally. It therefore becomes a matter of importance to state this principle precisely, so that we shall at least know what we are

* *The Philosophy of Religion versus The Philosophy of Science* pp. 169-70, 1935.

† *Philosophy and Logical Syntax*, pp. 15, 17-18, 1935.

doing when we bless or curse it. And in seeking to assess its value, we must avoid the error of ascribing greater or less weight to the opinion of a physicist than to that of another. Einstein is perfectly right in saying that if this principle holds for him as a physicist, it holds for him also if he is not a physicist ; and M. Maritain offers him a license which he cannot accept when he allows him to apply the principle in physics but not outside. If we talk nonsense and plead in extenuation that we are only talking as physicists, we fall short of the ideal of rational speech, distinguished precedent notwithstanding. The question must be considered on general rational grounds : are we to admit the unobservable into our scheme of things or are we not ? We must give an answer which is independent of the particular scheme which we want to uphold.

Let me for a moment adopt the legal method and state the case for each side independently, as an advocate might state it. Take first the case for the principle. It is necessary, say its supporters, as a safeguard against irresponsible invention. If we allow that an entity might exist and be significant to thought, although it is essentially unobservable, what is there to·prevent us from postulating any number of such entities and invoking them to remove any difficulty that might arise ? Suppose, for instance, I assert that there is a binkum sitting on the table in front of me, and that this tremendous fact, rightly understood, is the final, completely satisfying solution of the problem of evil. If you reject the principle in question, you have no grounds for denying the statement. You may say that you cannot detect my binkum, but I reply that of course you cannot, because he is unobservable. If you want to know how his existence solves the

problem of evil, I say that it is its nature to do so, and the definition of him, according to your own contention, is quite independent of any means you adopt to investigate him. If you ask, " What is a binkum, anyway ? ", I reply that that is immediately evident ; I cannot put it into words, but everyone knows what a binkum is. If you retort that you do not know, I shrug my shoulders and say that you must be speaking as a physicist.

Stupid as this example sounds, it contains a precise parallel to the case of simultaneity. M. Maritain and those who agree with him claim that the simultaneity of spatially separated events is something independent of our means of observing it. When asked what it is, they claim that its nature is immediately evident and that everyone knows it. But Einstein says he does not know it, and Einstein is an honourable man. And the only reply is that Einstein and his followers must be speaking as physicists.

We can now press home the point. We agree, let me assume, that science and philosophy are better without the binkum ; how, then, can we exclude him ? Only, says Counsel for the defence of the principle, by refusing to grant existence to anything that is essentially unobservable. Our only clue to reality is observation in one form or another. And if we admit that, we must go further. Not only must we deny existence to the unobservable : by the same token we are forced also to deny any unobservable property to an existent thing, otherwise we again open the door to the same abuses. And that means that everything whose existence we acknowledge must be definable ultimately in terms of observation—that for everything that we assert about it we must have, in the last

resort, some observable evidence.  The character
of the thing may, of course, be very different from
the character of the observation ;  a star, for
instance, may be a dazzlingly bright globe of gas
millions of miles in diameter, whereas we observe
only a faint pinpoint of light.  But there must be a
logical passage from every detail we assert about
the star to the observation of the pinpoint or other
observations of cognate character. If there is the
slightest relaxation of this requirement, in comes
the binkum with passport signed and sealed.

The case seems established, but we must hear
Counsel for the other side.   His argument is a
*reductio ad absurdum.*  Certainly we do not want
the binkum, he says, but your device for keeping
him out is both presumptuous and absurd ;  you
are throwing out the baby with the bath-water.
Consider for a moment what your principle implies.
It asserts that there is nothing in the universe
except what you can observe—nothing in the
physical world beyond the reach of your senses.
Suppose we were blind, like the men in Mr. Wells's
country ;  then, according to your principle, we
ought to act as they did, and deny existence to the
merely visible.  How do you know that the universe
does not contain things apprehensible only by
senses which you do not possess, which perhaps
you have lost in the course of evolution or have not
yet acquired ?  Moreover, what about the past ?
You cannot observe that, for it has gone ;  therefore,
you say, it has no meaning, it must not come into
our description of reality.  No flower has been born
to blush unseen ;  no feet trod the Earth before the
beginning of living memory.  Such nonsense neces-
sarily follows if your principle is granted.  Exclude
your binkum by all means—we hold no brief for

him ; but find some sensible way of doing it, without assuming potential omniscience.

The problem, I think, is now set clearly before us. If we admit the unobservable, there is no check on empty speculation : if we reject the unobservable, we confine the universe within the bounds of human potentialities and make nonsense of history. What are we to do ?

It is evident that we must begin by examining the word " unobservable." Both sides wish to exclude the binkum ; both wish to allow the universe a richer content than we can at present perceive. The " unobservable " that is to be proscribed must therefore include the obviously idle fancies but exclude the legitimately transcendental, and we must define it so that it does so inevitably, without special pleading. I need scarcely add that we must distinguish between " unobservable " and " unobserved "—between what we cannot possibly observe and what we have not in fact observed. This needs no elaboration ; we agree that an entity is not to be rejected merely because we have not observed it. But what do we mean by " possibly " observable ?

There are reasons of various kinds why we may be prevented from observing a thing, but I think they can all be summarized under four headings. First of all, there is what I will call the *practically* unobservable—i.e. that which is unobservable because of the practical difficulties of observing it. The far side of the Moon affords an example. That region is unobservable because we have not solved the problem of interplanetary travel, or succeeded in observing the Moon reflected in Mars, or learnt so much about the properties of matter that we are

able to deduce from observations of the near side of the Moon what the configuration of the far side must be, or performed some other feat of practical ingenuity. It is conceivable that in time this disability will be removed, so that practical unobservability may be merely a temporary characteristic.

Secondly, there is the *humanly* unobservable ; by which I mean the unobservable which is so because we do not possess the necessary faculties for observing it. I cannot, of course, by the nature of the case, give examples of this, but I can give analogies. A great deal of the universe would be humanly unobservable if we had no sense of sight ; and to the musically insensitive the significance of a great composer may be said to be humanly unobservable. If, then, there is in the universe some existence which no creature has the faculty of apprehending, that existence is humanly unobservable.

Thirdly, there is what I will call the *physically* unobservable. A thing may be said to be physically unobservable when we have the faculty for observing it if Nature will co-operate, but Nature gives that faculty no opportunity for exercise. Thus, if somewhere in space there occurred an event from which no signal—light ray or sound wave or anything else—proceeded to other places, and if there were repulsive forces which prevented us from ever reaching the place of occurrence, that event would be physically unobservable. We have eyes to see it, and hands to touch the bodies concerned in it, but we cannot use them for that purpose. An example important in physical history is the possibility of observing absolute motion. We have, in the Michelson-Morley experiment and similar

devices, instruments which are able to detect absolute motion if Nature will provide them with the necessary messengers, but this she fails to do. No matter how we may be changing our motion with respect to other bodies, our *absolute* motion always appears to be non-existent. (I am assuming, without prejudice, for the sake of illustration, that Professor Dayton Miller's experiments do not contradict this statement.)

Lastly, there is the *logically* unobservable ; namely, those things which we cannot claim to have observed without breaking the laws of reason. I doubt if this class is actual, since logic and observation are essentially independent, but, as we shall see, it must be included because a great deal has been written about it. An example might be the observation of an object both larger and smaller than a given object ; but I give this example with some hesitation because geometers have an uncanny knack of inventing spaces in which such relations might not be incompatible. Be that as it may, however, if we grant a certain minimum of common agreement—such as the acceptance of Euclidean geometry in the present instance—logical unobservability becomes an intelligible notion, and we will accept it as a candidate for inclusion in our principle.

Now this classification may be simplified ; for, whatever may be the ultimate truth of the matter, it is not necessary for our purpose to put the humanly and the physically unobservable in separate classes. I will therefore group them together and call them jointly the physically unobservable. The justification for this is that we cannot tell, in any given case, with which class we are dealing. If, for example, a certain substance appears taste-

less to everyone (i.e. its taste is unobservable), it is impossible to say whether that is because it has a taste which our senses are not keen enough to detect, or because it has no taste to be detected. There may be a distinction between the two cases, but if so it is beyond our apprehension. But now our principle is essentially one which, if valid, must be used ; it is not a creed which we are merely called upon to state and may then ignore. The humanly and the physically unobservable, then, become one class so far as our problem is concerned, for if in practice we reject one, we automatically reject the other also.

We have then three classes of unobservables, and I think the distinction between them may be expressed most simply in the following way. Let us suppose that we have discovered all the means of observation that exist in the universe, and know all their properties completely. We might then be able to imagine other means of observation which do not exist. Anything which would be observable by such imaginary means but not by the existing means, would be *physically* unobservable. ~ Anything which would be unobservable by *any* means, existing or imaginable, would be *logically* unobservable. Anything which would be observable by the existing means if we were also omnipotent, but which actually is unobservable because we cannot make full use of the means of observation which exist, would be *practically* unobservable.

We can now proceed a stage further. We have just seen that, for our purpose, the humanly and the physically unobservable become one class because we cannot at present distinguish between them. Let us look, then, at our latest classification, to make sure that the three types of unobservability

we have now reached are immediately distinguishable. It is fairly evident, I think, that they are, if we grant the initial supposition that we have discovered all existing means of observation. For brevity, I will call that the assumption of *omniscience*, and you will understand that by that word I do not mean knowledge of everything that exists or that can be observed, but complete knowledge of the existence and properties of every means by which observation is possible. For example, omniscience implies complete knowledge of all the properties of light, but not necessarily of all objects which are visible. Now, clearly, this assumption of omniscience is open to challenge, and it is therefore necessary to see how our classification looks if it is removed. Can we then still recognise the three classes as distinct ?

There is no difficulty, I think, with the logically unobservable. Since this class consists of things which are not even *imaginably* observable, it makes no difference how much we know of *possible* observability. There can be no possible means of observation that is not imaginable. There may be *unimagined* means, of course. We may make discoveries that take us by surprise, but those discoveries must have been imaginable, or we could not have apprehended them at all. Whatever we know of observability, then, does not affect our potentialities of observation, so that the logically unobservable is a definitely distinguishable class, independently of our assumption of omniscience.

The case is different, however, with the practically and the physically unobservable, for these classes *cannot* be distinguished if we do not regard ourselves as omniscient. We said that the far side of the Moon was practically unobservable, but if

we are not omniscient, how do we know that when we have overcome what seem to be the present difficulties, when we have made vessels with adequate air and food supply, and vehicles that we can drive accurately and swiftly enough to take us unharmed to the Moon during a week-end—how do we know that Nature will not then face us with some unexpected difficulty like that which she kept in store for our efforts to determine our motion through the ether ? If she does, and persists in doing the same kind of thing, we shall have to call the far side of the Moon physically and not only practically unobservable. And, on the other hand, when we say that absolute motion is physically unobservable, we are again assuming omniscience. We cannot observe such motion by optical, acoustical, electro-magnetic or any other means within our present knowledge, but without the assumption, who can say that there is not some undiscovered physical medium through which it may be detected ? If there is, absolute motion is merely practically, and not physically, unobservable. It is clear, I think, that unless we are omniscient the two classes are indistinguishable.

Our analysis of unobservability, then, finally brings us to this. If we assume that we are omniscient we can distinguish three classes—the *practically*, the *physically* and the *logically* unobservable. If we do not assume that we are omniscient we can distinguish only two classes— the *actually* and the *logically* unobservable, let us call them. The importance of this conclusion for our purpose is this, we are going to look at the actual practice of physics, to see what kinds of unobservable are excluded and what kinds are not.

If we find that a distinction is made between the practically and the physically unobservable, then we know that physics is assuming omniscience ; but if no distinction is made, then there is no such assumption.

Let me state the result at once, afterwards giving examples to justify the statement before proceeding to consider the validity of the principle we are considering in its definite form. The practice of physics is to recognise three classes—the practically, the physically and the logically unobservable. Of these it excludes the physically and the logically unobservable from its considerations, and aims at describing the universe in terms of the observable and the merely practically unobservable only. It thus assumes omniscience in the sense in which I have defined the word.

It will not take us long to see that physics includes the practically and excludes the logically unobservable. No physicist denies that the Moon has a far side in the same sense as it has a near side. We assume without question that the Earth has an interior, that there are stars outside the range of our telescopes, and regions beyond the obscuring clouds of the Milky Way. All these things could be observed if known means of observation have precisely the properties we believe them to have and we had the skill to make full use of them. Hence the practically unobservable is admitted to physical theory.

We may deal equally summarily with the logically unobservable. Nasty things have been said about the reasoning of some modern physicists when they step outside the bounds of their equations, but I do not think the bare, unadorned physical theories themselves have been called

illogical, either with pride or with shame. If, then, the structure of physical theory allows ontological significance to anything which is logically unobservable, it does so through an oversight, and theory will undoubtedly be reformed as soon as the oversight is discovered. We may say, therefore, that the practice of physics is to reject the logically unobservable.

But now, the unobservables whose rejection has caused all the controversy belong to neither of these classes. Let us fix our attention on the example of absolute simultaneity with which the discussion began. This, as we know, is rejected, and that cannot be because it is practically unobservable, for physics reeks of the practically unobservable. Nor is absolute simultaneity logically unobservable. We can conceive that the universe might be such that two events at different places might occur at the same time in an absolute sense, and that this fact might be observable. I should mention, however, that this has been denied, so I will presently show in more detail why absolute simultaneity cannot be regarded as logically unobservable, but for the moment I will take it to be established in order not to interrupt the main course of the argument. We cannot, then, escape from the conclusion that absolute simultaneity belongs to a third class of unobservables, which we shall see is what I have called the class of physically unobservables ; and the recognition of this class commits physics to the assumption of omniscience.

To see that absolute simultaneity is physically unobservable, let us look at the obstacle that prevents us from observing it. Why can we not say, in an absolute sense, that two events occurring at

different places are simultaneous ? It comes down to the fact that we can know of the events only through some agency which travels from them to us and takes time to do so. Consider two events— say the impact of a meteorite on a lunar crater and the outbreak of a new star in the Milky Way. We know of these events when we see them (or their effects) but we do not see them at the moments at which they occur because the light which makes them visible takes time to travel ; and although, perhaps, we observe them on the same evening, one may have occurred 100 years after the other. We can, of course, allow for this by calculating how long the light has taken to travel, but when we do so, according to standard methods and principles, we find that the results depend on the way in which we happen to be moving with respect to the bodies on which the events occurred. Further-more, we cannot distinguish in an absolute sense between one state of motion and another ; so far as we can determine, all states are equally valid or equally invalid. Hence we do not know what allowance to make for the time of travel of the light, and therefore cannot determine absolutely whether the events were simultaneous or not. The word " absolutely " is important. We can deter-mine quite definitely if the events are simultaneous if we assume that we are at rest or that we are moving in any definitely specified way ; but what we cannot do is to justify our assumption if it is challenged.

The unobservability of absolute simultaneity then, depends on the fact that we cannot determine unambiguously how long light takes to inform us of an event ; or, more generally, how long after an event it is possible for us to know of it. This

ignorance, of course, would not be necessary if we could know of an event immediately it occurred— if, this is to say, we could observe it by some messenger which travelled at an infinite speed. This is not pure fancy. Before the time of Romer, in the seventeenth century, it was believed that light might travel at an infinite speed, and before the theory of relativity arose, it was believed that gravitational action was transmitted instantaneously. It is not uncommon, too, to imagine that there might be instantaneous telepathic communication. We can at least, then, *conceive* that an instantaneous messenger might exist, and therefore absolute simultaneity is not *logically* unobservable. It is unobservable simply and solely because, so far as our present survey of the universe has gone, there is no evidence that it is possible to learn of a distant event at the moment at which it occurs. In other words, absolute simultaneity is *physically* unobservable.

It is important to emphasize this because, as I remarked just now, it has been claimed that absolute simultaneity is rejected because it is logically unobservable, and this claim has been made the basis of the philosophy of the logical positivists. Let us hear the late Prof. Moritz Schlick, one of the foremost and most able of the founders of this school of thought. He maintains that there are only two significant classes of unobservables, corresponding to what I have called the " actually " and the " logically " unobservable.

> " The distinction between impossibility of fact and impossibility of principle," he writes, " is absolute, without the slightest ambiguity ; it is not of such and such a degree ; it is essential."[*]

[*]*Erkenntnis* French translation in *Actualités Scientifiques et Industrielles*, No. 152, p. 25, 1934.

He then cites Einstein's rejection of absolute simultaneity as an example of the rejection of the *logically* unobservable, the " impossibility of principle." He makes it quite clear that he does not distinguish between the physically and the practically unobservable ; they are all included in the " impossibility of fact."

> " The statement, ' There are mountains 3,000 metres high on the far side of the Moon '," he proceeds, " is perfectly sensible, although our present technical skill is insufficient to assure us of its truth or falsity. *It would still be sensible if it were unquestionably established scientifically that man would never reach the far side of the Moon by any means.*" (My italics.) " The verification remains *conceivable ;* we are able to express what it would be necessary to do to decide the question, what it would be necessary to experience ; verification is *logically* possible, and that is all that matters."

I hope I have made it clear that if the *conceivability* of verification is all that matters, absolute simultaneity is verifiable and therefore not logically unobservable. We can conceive of the possibility of learning of an event at the moment of its occurrence. If, then, two separated events occur at the same instant, it is conceivable that we should observe them at the same instant. Now no one has denied that simultaneity of *observation* is significant ; it is only simultaneity of *occurrence* of the events that is in question. Hence, since simultaneity of occurrence is conceivably deducible from simultaneity of observation, it also would be significant if the logically unobservable alone were excluded.

We can now, at last, give a rigorous form to the principle which is the subject of our inquiry. It is this : that *only that which is practically observable*— i.e. only that which would be observable if we were able to use known means of observation to the

known limits of their possibilities—*is significant*. Our description of the universe must describe nothing else, must imply the existence of nothing else, must imply the *possible* existence of nothing else. If we do not accept this principle, we must reject relativity and a considerable part of quantum theory as worthless illusion.

The next step, clearly, is to examine the credentials of this principle on general rational grounds, but before doing so I want to give another example of its application, in order to emphasize the fact that it is not an unimportant appendage of physical theory but the very mainspring of the most prominent modern developments. Heisenberg's uncertainty principle is perhaps the best-known example, but I will not deal with that because it is too closely bound up with other factors which there is no time to consider. I choose instead an idea which stresses the point still more forcibly, because it is not generally regarded as exemplifying the principle in question but is attacked or defended on quite other grounds. The principle has taken root so deep in the minds of physicists that they employ it unconsciously, and justify their action by arguments which appear to others either incomprehensible or absurd. I am speaking of the idea that the physical universe is finite but boundless. This idea can be made intelligible in five minutes when presented as an example of the principle of rejection of the physically unobservable, and I believe that those who accept it are convinced of its rationality because they have already accepted that principle. They are not aware of this source of their conviction, however, and therefore have to justify their belief by saying that space is "curved," that it " bends back on itself "—an " idea " which

I do not think it is humanly possible to grasp except as a metaphor of the kind one meets with in the " metaphysical " poetry of the seventeenth century. Whether or not that psychological diagnosis is accurate, however, is unimportant ; the main point is to see that, in terms of our principle, the idea that space is finite and boundless is intelligible without calling on such unimaginable notions as curvature.

The idea is that if we were free to move about in space eternally, wherever Nature led, we would always find ourselves apparently in the midst of a collection of stars or nebulæ, though we could not for ever be meeting new objects, but would have to endure the tedium of seeing the old familiar faces endlessly, without relief. Like our former example, this idea, in its modern form, originated with Einstein ; let us see how it has impressed his contemporaries. Sir Arthur Eddington, who thinks very highly of it, writes thus :

> " Einstein made a slight amendment to his law to meet certain difficulties that he encountered in his theory. There was just one place where the theory did not seem to work properly, and that was—infinity. I think Einstein showed his greatness in the simple and drastic way in which he disposed of difficulties at infinity. He abolished infinity. He slightly altered his equations so as to make space at great distances bend round until it closed up."*

On the other hand, here is the stimulating Mr. Eagle again :

> " I fancy Einstein is honestly a bit tired and weary of ' curved space,' and is probably a bit sorry he ever suggested such a thing. But he did not know with what zeal other people would take it up and make a world-wide fuss about it. In pre-Einsteinian days if people had been told that an author's theme in his book had been that external reality only possessed three spatial dimensions, and that therefore

*The Expanding Universe, p. 21, 1933.

to talk of ' curved three dimensional space ' was pure non-sense, they would have replied that they thought only lunatics thought otherwise.  Now this conception is widely regarded in many circles as a ' probably may be true ' one.  This to me seems the measure to which both the scientific world and the general public have been bluffed by the theory.  Future scientific historians will probably regard the theory as a befitting product of a mad age in the world's history."*

Professor E. A. Milne is scarcely less contemptuous of this manifestation of Einstein's greatness.

" It is not necessary," he says, " to employ the vague, ill-understood, probably meaningless concept of closed, finite expanding space."†

Well, the symptoms are familiar.  Again we have the alternative estimates of supreme greatness and supreme folly, and we might suspect that the same misunderstanding is at the bottom of the trouble—as, in fact, it is.  Let us begin with a finite collection of stars (or nebulæ) in a space which we suppose extends to infinity in all directions (Fig. 1) ; we are somewhere in the midst of the collection—say at A. Now suppose we try to get outside into the empty spaces.  We cannot, because the gravitation of the system keeps all material bodies inside ;  the faster we move, the greater is the attraction, and we find ourselves following some such path as that shown in the diagram.  It is the same with light ;  that also is drawn back, and cannot illuminate whatever external objects there may be.  In fact, no physical existence that we know of can escape.  All that is perfectly conceivable ;  and I am not, of course, concerned with whether it is *true* or not, any more than I was concerned earlier with the truth of the statement that the velocity of light is finite ;  that is

*The Philosophy of Religion versus The Philosophy of Science, pp. 219-20.
† Zeits. f. Astrophysik Vol. vi, p. 83, 1933.

a matter for experiment, and my purpose is served if I can show that the idea of finite boundlessness *could* be true without violation of ordinary notions.
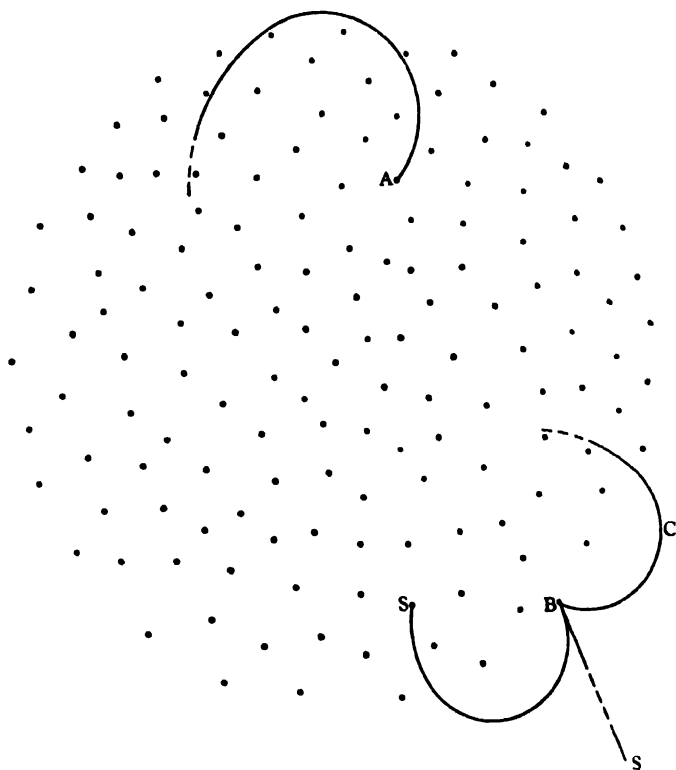


From " Through Science to Philosophy." By H. Dingle.
[*By Courtesy of the Clarendon Press, Oxford*
Fig. 1.

We picture, then, a finite universe from which we cannot escape. The next point is that it is impossible for us to know whether we are at its centre or its boundary ; in either case we appear to be surrounded by stars on all sides. Obviously an observer at the centre (or at A) sees stars all round,

but so does an observer at B, for the light of the
stars inside is bent, so that he sees them whichever
way he looks. The light of a star at S, for example,
is bent so that the star appears to be at S'. This is
a very ordinary phenomenon, exemplified every
time we look at a mirror ; an object appears in the
direction of the light that enters our eye, and not
necessarily where it actually is. We are therefore
not only confined within the universe, but neces-
sarily without the possibility of the experience of
being at the boundary ; however we move, and
wherever we go, we must always see the same kind
of thing—stars more or less uniformly distributed
all round.

All that is of the nature of simple, traditional
physics : now comes the crucial point. Since the
region outside the system is physically inaccessible
and unobservable, and the *experience* of being at its
frontier is physically unattainable so that it is never
possible for us to know whether we are there or not,
we leave these things outside our description of
the universe. We give the name " space " to the
volume which contains what we can observe, and
describe it simply according to our experience as
both finite and boundless. We can still, of course,
conceive that there is an infinite region outside
(wherever " outside " may be), but that is merely
another way of saying that the region is not
*logically* unobservable. The rejection of infinite
space is, in fact, precisely similar to the rejection of
absolute simultaneity : it represents an economy of
ideas—we are to introduce no conceptions which
are not necessary for the description of the physic-
ally observable.

It may be worth while to point out that this
account of finite boundlessness differs from the

statement that space is curved in the fact that it says nothing about any intrinsic *property* of space. There is no need to try to think of emptiness with a curvature ; we have simply to think of possible experience and keep within its bounds. Of course, for mathematical purposes the conception of curvature is useful because it allows us to employ the technique of Riemannian geometry to solve particular problems, and for the mathematician it may have some æsthetic value, as the poetry of Mr. T. S. Eliot is said to have for those " in the know." But for the purpose of understanding it is worse than superfluous ; it is definitely misleading.

What, now, of the validity of this principle, which has taken charge of physics and threatens to direct all future philosophical thought ? Let me repeat the principle ; it says that *nothing which is logically or physically unobservable is significant.* This statement must be appraised on pre-scientific, general rational grounds, and we can best approach the task by returning to the rival arguments set out earlier and considering them in the light of our rigorous statement.

The essence of the argument for the principle is that it is needed to prevent arbitrary invention of existences or arbitrary dogmas about them ; and it meets this requirement by saying, in effect, that everything that exists is observable by known means. The argument against the principle is that we have no right to assume omniscience ; that although it may be granted that the logically unobservable cannot exist, it is presumptuous to say the same of the physically unobservable. And as an example of the absurdity to which the principle leads us, it was pointed out that it would

require us to deny the reality of all that is past.

Now before coming to the main issue, it may be well to point out that this example does not hold good. The past is not physically unobservable but only practically so. Every event in the past inter- acted with the rest of the universe, and the effects of the interaction persist in one form or another ; it could therefore in principle be deduced from an analysis of those effects, just as the course of certain past events can be observed from a cine- matograph film. True, photography is a specially arranged type of interaction, the results of which can be analysed into a reproduction of the original events with relative ease, but that is merely a practical difference. Since radiation is going out from every event at every instant, and radiation, as energy, is indestructible, there always exists the physical possibility of tracing it back from what- ever form it may have assumed to its original character, and so recovering the past. Before con- cluding that the principle is inherently absurd, therefore, it is necessary to examine it with some care, for it leaves the possibility of a richer universe than might at first be suspected.

But making all due allowance for that, the criticism remains that the principle does cate- gorically deny existence to whatever cannot be physically observed, and that this implies an unwarrantable assumption of omniscience. We are therefore in this dilemma. If we deny the principle, we have no check on idle invention ; it may be that all that we know, and have taken such pains to find out, is trivial, while the great, important facts of the universe are not even suspected and are unattainable—at least with our present human faculties and knowledge, and per-

haps altogether. On the other hand, if we accept the principle, we close the door to all experience outside that which our present knowledge allows. Our mental, like our physical, universe becomes finite, and all that the future provides scope for is more ingenious manipulation of things we already know. And let me repeat that this dilemma is not a domestic affair for physicists. In physics it concerns at present only existences observable by sense perception, but clearly, it is equally relevant in the appropriate forms, to all spheres of thought in which we regard ourselves as apprehending some independent existence by means of human faculties.

This last sentence, I think, gives the clue to the solution of the problem : " all spheres of thought in which we regard ourselves as apprehending some independent existence by means of human faculties." That is the attitude which I have assumed throughout this discussion—the attitude of naive realism in which we picture an objective universe existing independently of our thought of it and our examination of it. It is the attitude which we always assume in everyday intercourse and in most scientific discussions also. We regard scientific research as an exploration of this independent universe, an attempt to discover what it contains and to understand the pre-established relation of one part with another. It is this conception of science or philosophy that makes possible the dilemma by which we are faced. If the universe exists independently of our experience of it, then clearly it is presumption on our part to deny that it can contain anything inaccessible to experience. And, on the other hand, if we abstain from this presumption we make all inquiry a mockery, for we have no guarantee that anything that we may

discover is more than a triviality, an insignificant part of a universe whose essential elements are eternally unknowable. We may, it is true, make a verbal escape from the dilemma by saying that, although the universe may be mainly unknowable, we proceed by an act of faith that we are given faculties capable of apprehending it completely ; but such an act of faith is both logically and practically indistinguishable from the assumption of omniscience—it is simply arrogance wearing the cloak of humility.

But suppose we take the idealistic view, regarding our experience, our observations, as the primary data, and the universe as a mental construct formed to give rational coherence to those observations. The whole matter then appears in a different light, in which the dilemma is no longer seen. The statement that nothing which is logically or physically unobservable is significant is simply a statement of our aim as scientists or philosophers ; it means that we confine ourselves to our purpose of deducing a universe from our observations and do not allow our fancies to intervene. There is no assumption of omniscience because there is no independent universe to know, and the arrogance disappears because we make no claim to know all the possibilities of observation. We set no limit to the possibilities of experience ; we simply refuse to assert anything for which we have no (direct or indirect) justification in experience, and as observation grows the universe grows also. The objection to the principle therefore vanishes completely, from the idealistic point of view.

On the other hand, the objection to denying the principle by no means vanishes. If we do not exclude the physically unobservable from our

description of the universe, we still have no grounds for not admitting the binkum and so reducing philosophy to a farce. When M. Maritain claims that a thing is independent of our observation of it, he immediately makes it impossible for us to know that we are saying anything of the least importance about it, no matter whether we adopt the realistic or the idealistic viewpoint. If we are realists the thing may be essentially beyond apprehension ; and if we are idealists we may form it equally legitimately from observation or from fancy.

The position, then, is this. If we take the realistic view, we are left with an unresolved dilemma, but if we take the idealistic view, the principle becomes simply a statement of the object at which science has aimed throughout its history. It is not my purpose to comment on the age-old problem of idealism *versus* realism ; I am concerned only with the attitude implied in modern scientific developments. And the point I want to emphasize as clearly and unmistakably as possible is that anyone who regards the recent trend of physics in general, and the theory of relativity in particular, as legitimate science or philosophy or intellectual activity bearing whatever name may be thought honourable, must either be an idealist or presume that he is omniscient. I do not wish in this place to plead the cause of either of the alternatives open to the humble. I am not anxious to cry " Vote for realism, and down with relativity ! " or " Support idealism and relativity, and throw realism to the dogs ! " That is a matter for personal predilection, but it is a matter for pure reason to show that those are the only alternatives open to anyone who is not prepared to assume that he is omniscient.

The striking divergence of opinion with which we opened can now, I think, be understood. Those who, like M. Maritain and Mr. Eagle, see the principle in question as an example of presumption arising from ignorance, are realists—by which I mean that they instinctively think as realists, whether or not they would accept the title. On the other hand, those physicists and philosophers who accept the principle are, by the same criterion, idealists, though most of them speak our ordinary, everyday language which has accommodated itself to the realistic outlook. In terms of that language their utterances necessarily appear arrogant ; what they apprehend instinctively as the wisdom of a self-imposed discipline is clothed in sentences which suggest to the realist the idea of arbitrary dogma. The divergence, arising as a difference of philosophical attitude, is accentuated by the necessity of expressing idealistic principles in realistic terms.

Let me, in conclusion, recapitulate the argument. The practice of modern physics is found on examination to imply that nothing must be included in our description of the universe that would not beobservable if we had full control of all known means of observation.   If we adopt the realistic view that the universe exists independently of our observation of it, this implies that the means of observation which we already know are sufficient to reveal everything that exists. We must in that case either make this assumption, for which we have no warrant, or else reject the recent developments of physics. If we choose the second alternative, we are left without any assurance that we can know anything of importance about what

exists outside us. On the other hand, if we adopt the idealistic view that the universe is constructed mentally by logical inference from experience, we see the implication of modern physics as simply the traditional thesis of science that the data we choose shall be pure experience, unalloyed with fancy or arbitrary dogma.

[H. D.]

SIR ROBERT ROBERTSON, K.B.E., M.A., F.R.S ,
*Treasurer and Vice-President, in the Chair*

IRVINE MASSON, M.B.E., D.Sc., F I C.
*Professor of Chemistry in the University of Durham*

# IODINE
## (I)

" A NEW and a very curious substance has lately occupied the attention of chemists at Paris.

This substance was accidentally discovered about two years ago by M. COURTOIS, a manufacturer of saltpetre at Paris. In his processes for procuring soda from the ashes of sea weeds, (cendres de vareck) he found the metallic vessels much corroded ; and in searching for the cause of this effect, he made the discovery. The substance is procured from the ashes, after the extraction of the carbonate of soda, with great facility, and merely by the action of sulphuric acid":—(*Experiment.*) ". . . the substance appears as a vapour of a beautiful violet colour, which condenses in crystals having the colour and the lustre of plumbago." (*Specimen*).

All that I have just said is dignified by marks of quotation ; I think the inverted commas will have become audible to you, for the words are those used by your own Humphry Davy when, 124 years ago next Friday, he sat down to write his first paper on the very theme of our discourse to-night. And surely, in this honoured place, which he first made one of the focal spots of chemical science, a discourse on Iodine could be introduced in no way more fitting, or more fortunate for the lecturer, than this.

It is not my purpose to recount the share which Humphry Davy had with the French chemists Clément and Gay-Lussac in fixing the elementary nature of iodine, its name and its chemical significance ; his share was a very important one, and although the story has been told more than once, and copied often, I am inclined to think that it would bear investigation anew, and that a scrutiny of *all* the available evidence would be found not to diminish, but to establish and to enhance the honourable distinction, the chemical insight, and the courtesy of the part which Davy played.

But I wish to take you forward from those times ; so far forward, indeed, as to take in our stride and accept nearly all that was done upon iodine by the nineteenth-century followers of Gay-Lussac, Davy and other pioneers. The body of material which is now at our disposal has many and very various aspects ; to try to survey the whole of it, flitting from one point of view to another, would be tedious for you even if it had been possible for me ; and so, having both the need to limit the field of discourse and the licence to do so, I shall treat to-night of two aspects alone.

## (II)

For our first view, let us take a wide scope ; let us see what part iodine plays in terrestrial economy, whether inorganic or organic or both. In this matter I shall muster facts drawn from many investigators, who have worked with very diverse ends in view.

The substance of our earth —its metallic core, the thick silicate sheath around this core, and the thin outer rind on which we have our being, and the ocean pools in that rind —all this earth is com-

posed as to 99.95 per cent. of its mass by some 14 of the chemical elements. Iodine is not one of these fourteen. It is, indeed, one of the very rarest of all non-metallic elements in the total make-up of the earth.

But although there is very little of it, that little is scattered almost everywhere. There is hardly any rock or mineral or water of the world in which delicate chemical testing does not discover *some* minute traces of iodine. Let me mention a figure. One ton of average rock contains about 300 milligrams of iodine. (*Specimen.*) This proportion is about the same, for example, as that of radium in pitchblende.

In the waters of the ocean the proportion of iodine is much smaller still ; a ton of sea water contains only 17 milligrams of iodine, as dissolved inorganic compounds. (*Specimen.*) The chemical reactions of geology have very rarely segregated any of these widely-dispersed traces ; there are scarcely any minerals containing iodine as an essential constituent, and the few that have been found are out-of-the-way mineralogical curiosities, small specimens rather than deposits or ores. Only in one place in the world have solid iodine compounds come together in quantity, and that is in the nitrate deposits of Chile, high up against the western slopes of the Andes. There the element occurs as inorganic iodate and there are uncounted thousands of tons of it. Yet even so, the iodate is not by itself in veins or beds, but is simply one of the minor impurities in the great accumulations of sodium nitrate. We might be able to explain how the iodate got there if we knew how the nitrate had got there ; but, in spite of much interesting work, all that can safely be said about the origin of Chile

nitrate is that no single hypothesis has yet met all the known facts. What is indubitably true is that the world's commercial supply of iodine (with trifling exceptions) comes from the Chile nitrate factories, from which it is exported at the rate of nearly 1000 tons of the element yearly.

Now geology has joined with chemistry, physics and subatomic science to learn with very fair certainty that this earth has been as it were a crucible for chemical reactions for about 2000 million years. And what I have so far mentioned would suggest to you that in all these long chemical-geological processes, iodine played a part that was practically nothing. Certainly we have learnt to beware of stigmatizing an element as geophysically negligible merely because there is only a little of it—a lesson which the case of radium has taught us —but iodine is not radioactive like radium ; it lends nothing to the earth's supply of heat ; and in view of its great sparseness, it is therefore true to say that had no such element as iodine been created, the inorganic earth would not be appreciably different from what it is.

But now notice a fresh factor.

Somewhere during that vast stretch of time— certainly 800 million years ago, possibly much earlier—organic life began ; and with that, iodine began to matter. First it became not wholly dispersed, but partly to be segregated from solution in the sea into the colloids of living organisms floating there. Then, as the forms of living things developed, the iodine was not only taken up into them but became a chemical part of some of them. They developed still further and more variously ; and now the vertebrates, and we ourselves, could not continue without iodine, either as individuals from

month to month or as species from generation to generation.

Let me elaborate this ; for the biochemistry of iodine contains facts that are, I believe, worth thinking over from points of view beyond the purely chemical.

Many analyses have been made in different parts of the world, in which have been measured the traces of iodine in a rather wide variety of living materials, both animal and vegetable, dwelling on land, in fresh water and in the ocean. All the data which are definite go to show this : if for the moment we exclude the vertebrate animals and their organs, three living creatures stand out far beyond the rest as being rich in iodine. These three are seaweeds (brown and red), bath sponges and certain classes of corals. One of these is vegetable, two are animal, but all three, you notice, are marine ; and this is so notwithstanding the fact that the sea-water is excessively poor in iodine compared (for instance) with many soils and most rocks. The iodine content of these organisms is thousands or even millions of times more than in the sea-water in which they live ; it may reach one-half per cent. of the weight of the dry tissue in seaweed, one per cent. or more in the sponges, and in a coral as much as eight per cent.

And if now we bring in the vertebrates, which we had momentarily disregarded, we find that a vertebrate animal—a fish, an ox, or a human being —is very poor in iodine as a whole, but one small part of him is quite rich, namely his thyroid gland. Thyroid tissue is about as rich in iodine, weight for weight, as seaweed.

We naturally ask, in what chemical form is this iodine present ? As to seaweed-iodine, chemists

have not yet succeeded in deciding how far it may be there as inorganic substances, how far as organic, nor have they yet identified any organic compound of which it forms a part in the vegetable protein of the seaweed. But as to the sponges and the corals we have definite knowledge. The surprising discovery was made a generation ago, by the work of Drechsel (1896) and of Wheeler and his associates (1905–07), confirmed by that of Mörner (1907–13), that the skeletal protein both in the bath sponges and in horny corals releases its iodine as the definite organic chemical compound of which here is a specimen and here is the molecular structure :



(3 5 di- iodo- tyrosine)
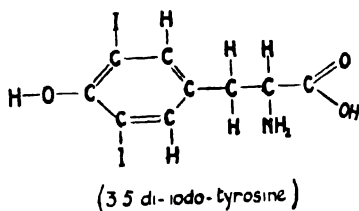
Notice the two iodine atoms in the molecule, attached to the aromatic benzene ring.

Before saying anything else about this substance, I should like to lay stress upon the point, and to ask you to keep it in mind, that in animal and vegetable tissues the presence of iodine, or indeed of its chemical brethren bromine and chlorine, as part of a carbon compound is a thing excessively rare. In laboratory organic chemicals it is familiar, but in natural organic substances it is almost describable as a freak occurrence. Yet we shall find presently that it is a freak we could not do without.

In view of the rarity of this occurrence, let us see by what means our compound has been formed

within these special animal tissues. We can first point to its very close relationship with a chemically familiar and frequent constituent of proteins, namely, tyrosine. Here is tyrosine, and its molecular formula is—
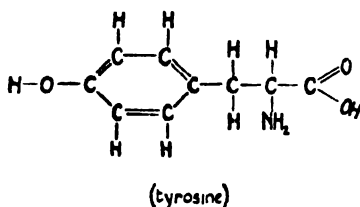


(tyrosine)

Our substance is di-iodo-tyrosine. Iodine atoms have replaced hydrogen atoms in the aromatic ring, as the working model shows. (*Model.*) In the laboratory it is quite easy to iodinate tyrosine so as to make this compound, as Oswald showed (1909) and as Bauer and Strauss have lately (1935) learnt in more detail. But the special reagent used for this purpose in the laboratory, namely, iodine in the free elementary state, is not available to the sponge or the coral, the living animal has to make its di-iodo-tyrosine from protein material and the inorganic saline *iodide* which is present in traces in sea-water ; and to do that, the participation of an oxidizing agent of some kind would be necessary. It is reasonable to suppose—though direct experiments are, I think, lacking—that some oxidase or peroxidase enzyme, such as are ordinarily present in living creatures which use air, has been responsible in the sponge or coral for turning tyrosine and iodide into di-iodo-tyrosine. There is fairly sound evidence to warrant us in doubting that the sponge or coral gets its di-iodo-tyrosine ready-made in its plant food ; it may get the sub-

stance of tyrosine from what it eats, but apparently the iodination of that tyrosine goes on in its own tissue, the iodine for the synthesis being drawn by oxidation from the absurdly dilute solution of inorganic iodide in sea water.

Two close parallels to this curious chemical occurrence are worth our notice : both of them marine also. From the protein of a horny sponge, Mörner (1913) isolated di-bromo-tyrosine. The biochemical mechanism of its formation must be almost exactly the same as that of di-iodo-tyrosine. The other kindred example is in Friedländer's interesting discovery (1909, 1922) that the classical Tyrian Purple, the Imperial Purple of the ancients, manufactured from the glands of the shellfish Murex or Purpura, is—quite unexpectedly— di-bromo-indigo. Of natural organic compounds of the allied but far commoner element chlorine, none was known until three years ago, when the study of lichens, at Dublin by Nolan and at Geneva by Pfau, has disclosed in them several aromatic compounds rich in chlorine ; and lately also Raistrick has shown that certain moulds can convert inorganic chloride into organic chlorine compounds.

Returning to di-iodo-tyrosine, we should like to know whether it—or, speaking more generally, iodine—is necessary to these marine creatures, vital to their existence, or whether it is merely a by-product, an incident in their biochemistry : would they, for example, flourish as well, or better, or worse, in sea-water artificially dosed with controlled amounts of iodide, or deprived of it altogether? There seem to be no such experiments on record, and the biologists whom I have asked therefore naturally tell me they do not know. It is

worth noticing, however, that the iodine compound is accumulated not in the fleshy cell-tissue of the sponge or the coral, but in the horny skeleton on which these living cells sit ; and this skeleton, though mechanically useful to the cell colony, may almost be called an excretory product of that colony. It is for the animal physiologists to decide whether in these cases the skeletal protein is serving as a mere dump for di-iodo-tyrosine that has been discarded by the cells as useless, or whether on the contrary it acts as a reserve store upon which the cells can draw for part of some vital process. This question, unsettled though it is, is not without significance in the light of the bio-chemistry of animals higher up in the evolutionary scale, to which we now turn.

It is a very wide jump from sponges and corals to these higher animals. For, as we learn from biologists, sponges and corals are extraordinarily low in the scale : segregations of animal cells, with not much differentiation and with no co-ordination of parts ; and their fossil history reaches far back into Cambrian times, long before any forms of land life had evolved. These creatures, then, already have the power of fixing inorganic iodine as organic di-iodo-tyrosine ; let us go from them right up the biological scale to the vertebrates, and among vertebrates straight to man, and there briefly examine the fixation and the function of iodine.

In a man's body the amount of iodine is of the order of twenty or thirty milligrams. But about half of this is lodged in his thyroid gland, the rest is merely incidentally scattered. If a man is given a small dose of iodide, his thyroid seizes a share of it which is disproportionately great in relation to the small size of the gland.

Next, it is nowadays well known that the thyroid makes, stores, and releases a hormone which controls the basal metabolism of the body. It is also well established that irregularity of thyroid activity is connected with inadequate traces of iodine in the food ; and further, that such irregularity is responsible for the common disease, goitre ; that a dearth of the hormone impedes full growth in young people, more dearth causes myxœdema, and extreme deficiency at early stages causes that most distressing failure in mental and bodily development, cretinism. That the thyroid hormone is vital as a factor in the health of individuals, nations, and humanity is perfectly evident ; the hormone is indispensable.

This is true not only among human beings, but among all kinds of vertebrate animals. And how potent the thyroid hormone is in other ways is illustrated by the fact, first found by Gudernatsch in 1912, that it provokes the metamorphosis of amphibian larvæ. Tadpoles, in fact, turn into frogs much more quickly if given doses of thyroid extract. And through the kindness of my friend Professor D. M. S. Watson, of University College, London, we can see here a Mexican axolotl, fated by lack of thyroid hormone to live and breed and die in axolotldom (but quite happy !), and here one of its brethren. This brother may almost be said to have been magicked, for a single minute dose of thyroid was given it when it was an axolotl, and it has turned into this admirable salamander.

The known correlation between iodine in food supplies and the proper working of human and animal thyroid glands was given a rational basis when, nearly twenty years ago, Kendall in Chicago succeeded in extracting the hormone itself and

showed it to be a crystalline organic compound of iodine. He gave it the name thyroxine. And it is to the brilliant work of Professor C. R. Harington, of University College Hospital, London, that we owe our knowledge of what thyroxine really is. From what he and his associates have published during the last dozen years I merely pick out one or two leading facts which bear directly on our train of thought.

Harington proved that the molecule of the hormone thyroxine is
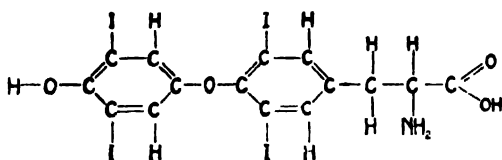


THYROXINE

Postpone for a moment any close examination of this rather complex structure, but notice the four iodine atoms, attached to the two aromatic rings ; and mark also the fact—as it seems to be—that the presence of these iodine atoms, attached at those points, is essential to the physiological potency of the substance ;· no other atoms will serve in their places. A teleologist might well find in this hormone molecule the sole *raison d'être* of the element iodine.

Harington and Barger synthesized the compound artificially. Here is a specimen of synthetic thyroxine which we owe to the courtesy of Dr. F. H. Carr, of British Drug Houses, Ltd., and Professor Harington has himself been so kind as to let me show you one of his own microphotographs of the crystals of the material, synthetic *l*-thyroxine.

Now Harington further proved that in the

thyroid gland the iodine is present not only as thyroxine, but also as the compound which we have already met, di-iodo-tyrosine. Both are chemical components of the highly complex thyroid protein. The di-iodo-tyrosine component appears to have little *direct* physiological activity of its own, compared with the thyroxine. But its significance is that with high probability it is to be looked upon as the precursor of thyroxine in the sequence of chemical processes for which the gland is responsible. Our model shows, necessarily in a much simplified fashion, how the two are related. Here are two molecules of di-iodo-tyrosine ; if some chemical agency could lop off the tail of one molecule (and in other branches of protein chemistry such agencies are known), you can see from the working of the model how the residues could couple together so as to form the more complex thyroxine.

This synthesis has, I understand, not been accomplished in the biochemist's laboratory at present ; Harington's pioneering synthesis of thyroxine had to use methods and substances not natural to living tissue ; but meanwhile, on indirect evidence, and in the absence of proof to the contrary, it is scarcely possible to avoid the view that the thyroid achieves its synthesis of thyroxine *via* the intermediate stage of di-iodo-tyrosine.

. . . . .

So let us sum up and take stock of what we have learned. The opportunity for the natural fixation of inorganic iodine into an organic form has in all ages been extremely slight ; not because the process is intrinsically difficult (for it is not), but because the supply of inorganic iodine has always

been extremely scanty. Accordingly iodine is not one of those elements which, like sulphur or phosphorus, became normal ingredients of common proteins. Living things had next to no iodine to draw upon, so most of them managed and evolved without it. But very early in the history of multi-cellular animals in the sea—if we may make an inference from those simple forms which have survived into this age—the organic fixation and accumulation of iodine was accomplished by some of them, and what they did with it was to make it into di-iodo-tyrosine.

This same faculty of making di-iodo-tyrosine, we find, has been preserved, or at any rate repeated, by the vertebrate animals, localized in their thyroid glands. And besides that faculty we have put our finger upon one further chemical step, which has apparently accompanied the evolution of the vertebrates, is characteristic of them, and is indispensable to their being—namely, the final conversion of di-iodo-tyrosine into thyroxine.

### (III)

Now having tried to survey with you the importance of iodine as a biochemical element, I wish to look with you at another aspect of its organic chemistry, of its compounds with carbon. We leave biochemical considerations, and pass to open certain chapters of a story of which little but the headings have as yet been written ; some of it not new, some of it new, to the development of which at Durham my research students, Dr. Race and Dr. Pounder, contributed in a way which I am glad to take this opportunity of acknowledging.

If we look at that part of the Periodic Classification which contains iodine (*Diagram*), we see

iodine in the same group as fluorine, chlorine, and bromine : the halogen group. And indeed, a great many of its compounds, both inorganic and organic, differ from the corresponding compounds of fluorine, chlorine, or bromine only in degree, not in kind. In them iodine behaves much as we should predict of it as a mere extension of the halogen family. But while it is true that iodine in the Classification falls at the foot of a group, a column, it is also true that it occurs at the end of a series. a row ; and so, especially when the iodine atom has directly linked to it more than one other atom—in short, when it is multivalent—the chemistry of iodine is not so much that of a halogen element as like that of its side-neighbour tellurium, and, still more closely, antimony and its congener arsenic. Chemists who bear in mind these connexions will be saved from thinking of iodine as the black sheep of the halogen family, particularly in those parts of its chemistry of which I wish to put before you a few connected examples.

A starting point for this branch of chemistry is from the iodine derivatives of benzene, or more generally, the aromatic iodo-compounds. More than fifty years ago, Willgerodt (1885) observed that the liquid iodobenzene will unite directly with elementary chlorine, forming a yellow crystalline compound. (*Experiment.*) In this compound each molecule of iodobenzene is joined with one molecule of chlorine. It is not very stable and cannot be stored for long without decomposition. The mere fact that such a complex exists would not itself matter much ; for there are hundreds of complexes between two substances, in which the chemical behaviour of the complex is simply that of a mixture of its components. So, indeed,

Willgerodt may himself have regarded this sub-
stance, for he left it unstudied for seven years.
Meanwhile, however, that fine chemist Victor
Meyer, during investigations in 1892 which we
have not time to examine, discovered a compound
of a type hitherto unknown to chemistry, and
named by him iodoso-benzoic acid.

This prompted Willgerodt to resume the study of
his own compound ; and by treating it with a
solution of alkali, he obtained for the first time
iodosobenzene.

But as we cannot well see this done here with the
substances themselves, let us follow what happens
by watching the working of this model.

Here is benzene, the ultimate parent body,
modelled approximately to scale—$C_6H_6$.　　By
replacing one of its hydrogen atoms by an iodine
atom we get iodobenzene, $C_6H_5I$. We now bring
into reaction with this (as we did just now in actual
fact) molecular chlorine, and the twin chlorine
atoms proceed to attach themselves to the iodine
atom of the substance. Next, we treat the product
with alkali, that is, with a watery solution contain-
ing hydroxyl ions. Here are two hydroxyl groups,
OH, and each· has a bright blue light on it to
symbolize the negative electric charge which
makes the group that carries it an ion. These move
up to the attack : the chlorines, being atoms which
prefer the ionic state to any other, rob the hydroxyls
of their electrical charges and go off as chloride ions,
their places on the iodine atom being simultane-
ously taken by the hydroxyl groups. And from
this product, next, there is spontaneously released
a molecule of water, $H_2O$, leaving as the final
product the substance iodosobenzene, $C_6H_5IO$.

As a result of Willgerodt's extensive work, it

appears that the majority of aromatic iodo-compounds will perform just such reactions, and yield iodoso-compounds, the essential feature of which is the group —I : O.

The salient properties of iodoso-compounds are these : they are cream-coloured powders, which do not melt without decomposing ; they are not volatile, yet they have a smell, rather unpleasantly noticeable, and they are rather poisonous. They do not keep very well, hence I have not brought any specimens. They do not dissolve in water or any of those solvents which would facilitate a chemical study of them. Nor do they dissolve in alkalis. But they do unite with acids, forming salts, such as the crystalline acetate. (*Specimen.*) So iodoso-benzene is a kind of base.

But the outstanding chemical character of the iodoso-group is the very loose attachment between its iodine atom and the single oxygen atom. For this reason iodosobenzene is a particularly strong oxidising agent towards many substances; and in so acting, it is reduced again to iodobenzene, so that the iodine atom reverts from trivalency to univalency.

Here, however, we come upon something with which chemists are very familiar among the *inorganic* compounds of the halogen elements ; namely, that their oxygen compounds decompose so that the halogen atoms abandon an inter-mediate valency in such a way that some of them assume a higher valency while simultaneously the rest assume a lower. And iodine retains this tendency in these *organic* compounds. Accord-ingly, if iodosobenzene is not offered some reagent which it would oxidise, eventually it oxidises itself.

$$2\ C_6H_5IO \longrightarrow C_6H_5IO_2 + C_6H_5I$$
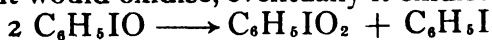
Iodine valency (3)　　　　　(5)　　　　(1)

Half of the molecules get all the oyxgen, the other half are left as iodobenzene with none ; so we get a new compound richer in oxygen than iodoso-benzene, namely, iodoxybenzene.

What is the new substance, iodoxybenzene ? This pure white crystalline substance (*Specimen*), scarcely soluble in any ordinary solvent, has commonly been regarded, during the forty or more years since its discovery, as an unprofitable body, whose only obvious properties were that it explodes if it is heated much over 200°, and that it, like iodosobenzene, can rather easily be deprived of its oxygen and revert to iodobenzene. Actually iodoxybenzene, as we have found at Durham in the course of our work on iodine, has much more personality than that, and teaches us a good dea of general chemistry ; let us see something of this.

The formula recalls nitrobenzene—$C_6H_5IO_2$, $C_6H_5NO_2$—and there are indeed ways in which the two have a real chemical similarity, different though they look physically. (*Specimen.*) Here are models of them, side-by-side, approximately to scale. In both molecules we are fairly sure, on various grounds, that the nitrogen atom or the iodine atom, as the case may be, is electrically positive with respect to the oxygen atoms, owing to the existence of co-ordinate linkages within the molecule. In the models the dull red glow on the iodine and on the nitrogen symbolizes this positive polarity, as against the negative polarity which is imparted to the oxygen atoms and which is here symbolized by a dull blue glow on them. Follow out some consequences of this polarity.

First notice one marked structural difference between iodoxybenzene and nitrobenzene ; the iodine atom is large—indeed, it is larger than any
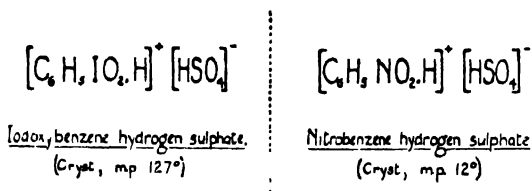
other non-metallic atom—whereas the nitrogen atom is quite small. Hence the nitrogen atom is screened off, as it were, by its attached atoms, from outside agents, whereas the large iodine atom is much more exposed. One might almost say that the nitrobenzene molecule presents to the outside world a surface whose electrification is practically all negative, whereas the iodoxybenzene molecule offers both its negative and its positive poles to any outside agents that may be electrically attracted by either the one pole or the other.

This very strongly *di*polar character of the iodoxy-group probably helps us to understand why iodoxy-compounds are rather refractory insoluble solids ; for their molecules could cling together by strong electrical dipole attractions. There is a need here for proper X-ray analysis, for nothing has been done.

It is in the chemistry, however, of the iodoxy-compounds that the dipole shows its effects most plainly.

First, the oxygen atoms, being negative, can attract and hold the positive hydrogen ions of strong acids, so as to yield salts. This applies, as might be expected, not only to iodoxybenzene but to nitrobenzene as well, and in both cases we have isolated, for instance, crystalline salts with sulphuric acid :—

$$[C_6H_5IO_2.H]^+ [HSO_4]^-$$

Iodoxybenzene hydrogen sulphate.
(Cryst, mp 127°)

$$[C_6H_5NO_2.H]^+ [HSO_4]^-$$

Nitrobenzene hydrogen sulphate
(Cryst, mp 12°)

The models show you this, in action.

Looking now at the other side of the atomic

dipole, we find that iodoxybenzene also allows the opposite sort of salts to be formed ; that is to say, salts with alkalis. Reluctant though iodoxybenzene is to dissolve in ordinary fluids, it does dissolve in any alkali, and it has been proved to do so, because the negative hydroxyl ion of the alkali is attracted to attach itself to the positive iodine atom in the molecule, as the working of the next model shows. The chemical equation is also shown on the screen.

$$C_6H_5IO_2 + Na(OH)^- \longrightarrow Na[C_6H_5IO_3H]^-$$

| Iodoxy-benzene | + | Sodium hydroxide | = | Sodium phenyl-iodoxylate |
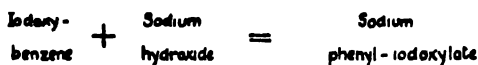
We call the salts so formed with alkalis " iodo-xylates " ; those given by iodoxybenzene are the phenyl iodoxylates. It is to be noticed that nitro-benzene gives no such salts, and we may put this down, as I indicated before, to its positive nitrogen being so small as to be screened by the benzene ring and by the negative oxygen atoms, which ward off an approaching hydroxyl ion. But the elements arsenic and phosphorus, whose atoms are larger than those of nitrogen, do yield just such salts, which are known as the organic arsonates and phosphonates.

The iodoxylates exist only in solution in alkalis ; and you have to work quickly in order to discover their existence at all, and still more quickly to determine their exact formulæ and nature ; for in a very few minutes the dissolved iodoxylate spon-taneously turns into something else, even at the ordinary temperature. If it is heated, it breaks down in a way which I wish to show you by means of this model. (*Model.*)
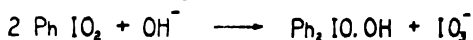
Here is the phenyliodoxylate ion, with a bright

blue light on it to symbolize the negative ionic charge which the whole complex carries. When its solution is boiled, this complex breaks up, thus, emitting the stable inorganic iodate negative ion, $IO_3^-$, which carries off all the ionic charge and leaves behind simply benzene.

$$C_6H_5.IO_3H^- \quad \text{---------} \quad IO_3^- \ + \ C_6H_6$$
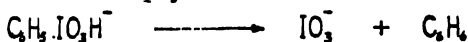
phenyl-iodoxylate  ▬  iodate  +  benzene

This elimination of organic iodine in the inorganic form of iodate turns out to be very characteristic of the substances we are discussing, and I am going to show you two more examples of it and of its consequences. Here is a second instance of its happening. If an alkaline solution of iodoxybenzene, containing iodoxylate, is left to itself at the ordinary temperature, the iodoxylate quickly undergoes a change different from that which we have just examined. The organic product is no longer benzene, but a body of a new type ; the ruling condition which determines its formation is, however, the same as before, namely, the expulsion of the iodate ion. As the working model* will show us, two molecules are now concerned instead of only one :—

$$2 \ Ph \ IO_2 \ + \ OH^- \quad \text{----} \quad Ph_2 \ IO.OH \ + \ IO_3^-$$

Iodoxy-      Diphenyl-iodyl
benzene      hydroxide  + iodate

*FIG. 1 shows drawings of four successive stages of model 5, which also illustrate the general nature of the other models. The backgrounds were black. Positive ions were marked by the emission of a bright red light from the group of atoms ; negative ions by that of a bright blue light (indicated in FIG. 1). Polarities within molecules were shown by a red glow over the disc of a positive (donor) atom and a blue glow over the disc of a negative (acceptor) atom, as suggested in the drawing of the molecule of iodoxybenzene on the right in the first stage.
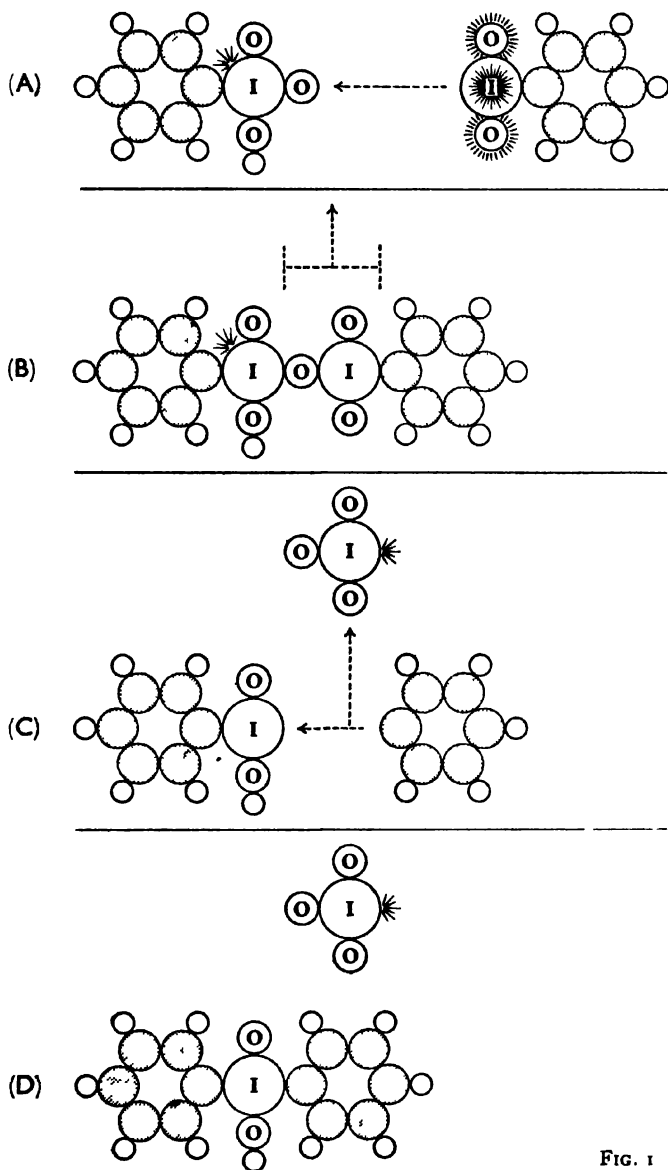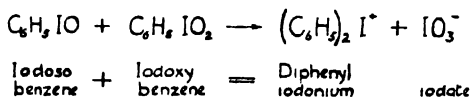
FIG. 1

All that I need say just now about these diaryl-iodyl compounds (*specimen*) is that they have their analogies outside the chemistry of iodine, in the diarylarsinic and -phosphinic acids, and that they have also many interesting characters of their own.

Our final reaction is not a new one, but it again illustrates the principle established by the two we have just seen. If iodoxybenzene is treated in alkali with iod*oso*benzene, these two kinds of molecule conspire, as it were, to produce between them an iodate ion. How this happens, and what else is formed as a result, will be seen from our last model, which, like all the others, owes its merits of construction to the inventiveness of Mr. Beecroft.

You see how, by the extrusion of a negative iodate ion from the conjunction of the two neutral molecules, there is left a positive ionic complex, in which both aromatic groups are tied to a single iodine atom. This positive ion is the diphenyliodonium ion.

$$C_6H_5 IO + C_6H_5 IO_2 \longrightarrow (C_6H_5)_2 I^+ + IO_3^-$$

$$\underset{\text{Iodoso benzene}}{} + \underset{\text{Iodoxy benzene}}{} = \underset{\text{Diphenyl iodonium}}{} \quad \underset{\text{iodate}}{}$$

Its discovery was due to Victor Meyer and Hartmann in 1894, and the two short papers in which they described the work show well Victor Meyer's range of vision and the quickness with which he profited by an unexpected observation that upset all his initial ideas, in the way that marks a true chemical explorer. The existence of the iodonium compounds shows iodine as one of a class of elements to which no other halogen elements belong, but which contains most other non-metallic elements and some of the metallic ; the

ammonium salts from nitrogen, the sulphonium from sulphur, and the iodinium from iodine, are part of the set shown in the table (*Diagram*).

The hydroxide of iodonium is a strong alkali, as in all the analogous cases shown in the table ; it forms crystalline salts with any acid, just as ammonium hydroxide does. In some of these salts, as Victor Meyer and Hartmann showed, the iodonium radicle is strangely reminiscent of thallous or of lead salts, as in the precipitation of the chloride, the bromide, the iodide, and the chromate. (*Experiment.*)

In all these cases, then, we have seen the organic complex, whether it is itself an ion, a mole- cule, or a pair of them, ejecting an inorganic iodate ion, usually in an alkaline medium, and recompos- ing what is left into a new organic complex. It has seemed to me that the converse process, the return of iodine from inorganic iodate into the organically combined state, would be interesting ; it might even, conceivably, touch the biochemical problem of the iodine cycle in Nature ; but whether so or not, we find in fact that under conditions of high acidity it proves a very easy matter in the labora- tory to convert inorganic iodate directly to organic compounds of multivalent iodine, and thence, if desired, to iodo-compounds. More than this need not be said here.

If what I have dealt with in the second half of this discourse has appeared rather narrowly speci- fic, let me remind you that each of the *benzene* compounds I have touched upon is itself a parent type of many derivatives ; and further, even apart from this mere multiplicity, the general chemistry of the organic compounds of oxidized iodine

presents a large field of chemical interest. And this fact brings us back to what I have tried to make plain in both parts of my discourse, namely, that we must think of iodine as an element which, like nitrogen, is not inorganic alone, but is emphatically an organic element also.

[I. M.]

SIR ROBERT ROBERTSON, K.B.E., M.A., F.R.S.
*Treasurer and Vice-President, in the Chair*

SIR WILLIAM BRAGG, O.M., K.B.E., M.A., D.Sc.,
Pres.R.S., M.R.I.
*Fullerian Professor of Chemistry*

# ICE

In the last few years the examination of the crystal-line character of ice and snow has revealed certain facts which suggest to us new explanations of old problems. For ages the movements of ice and snow on the mountains have, of course, been a source of wonder and awe. In the middle years of the nineteenth century there was an organised attempt to measure the stately movements of Alpine glaciers and to determine their character. It was quickly established that the ice flowed down the valleys like water, bending and dipping as the character of the ground required. Agassiz and Forbes showed independently in 1842 that the centre of the glacier moved faster than the edge. Tyndall in 1857 continued this work, and described vividly in his book on *Glaciers*, from which the illustrations on Plate 1 are taken, the staking of a line across the *Mer de Glace*, and the careful measurement of its subsequent position. This curious plasticity of a material so brittle as ice was at that time a great difficulty, and provoked attempts at explanation to which I will presently refer.

Later in the century it became the custom to frequent the Alps in the winter for the sake of the various forms of sport which they afforded. It was then found that ski-running and the other sports

were greatly affected by varieties in the character
of snow. This was a new incentive to the study of
ice forms. The members of this Institution will
remember the most interesting discourse on the
*Nature of Snow* which Dr. Seligman gave here in
January of last year. He dealt particularly with the
properties of snow which he, an expert ski-runner
and a keen observer, had studied with so much
attention and ability. Incidentally his mountain
pictures were so grand and impressive that I have
asked him to allow me to show a few of them once
more on the screen. Two of these are reproduced in
Plates II and III. I am grateful for the permission
to do so because the relation between these 'mag-
nitudes and the minute details of the water mole-
cule and the ice crystal is itself an occasion for
wonder.

The beautiful regularity of the snow flake, es-
pecially as it is found in some northern countries,
was also a subject of interest to early observers; but
it was only towards the close of the last century
that series of observations were made and classi-
fied. It has been an occasion for perplexity that the
six rays of any one flake are so like each other, and
yet that no two flakes agree perfectly. Examples
from the collection made by Bentley in U.S.A. have
often been shown in this theatre. Quite lately a
fresh series of observations made in Japan by
D. Nakaya reveal new beauties, and I would like
presently to show examples of these. (Plate IV.)

Lastly, an arrangement of atoms and molecules
in ice and snow is a subject of extraordinary interest
to the chemist and physicist. Oxygen forms half
the known material of the earth's surface. Its com-
binations with hydrogen to form water ($H_2O$) or
hydroxyl (HO) are fundamental to a vast series of
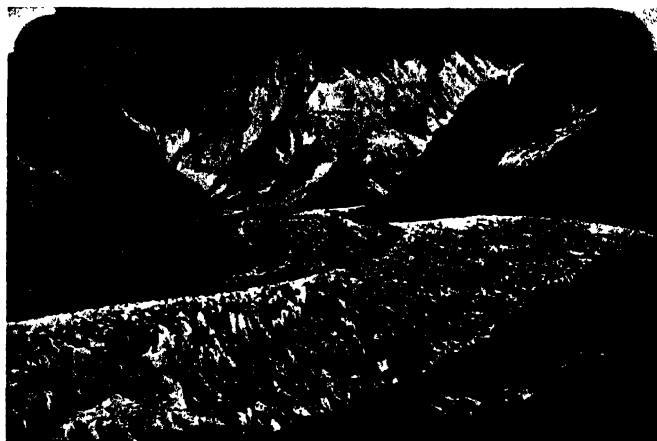
A. From the frontispiece of Tyndall's "The Glaciers of the Alps"; the Mer de Glace on which he observed the movement of the ice in 1857.
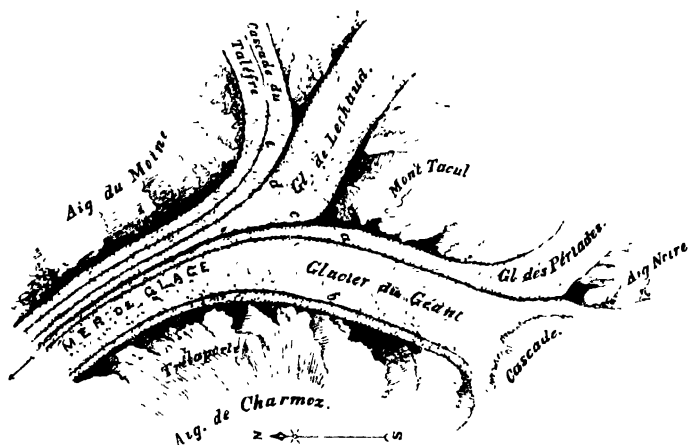


B. A diagrammatic sketch from Tyndall's "The Glaciers of the Alps" showing the river-like movement of a glacier—the Mer de Glace.

PLATE I.

processes in Nature. The bonds by which the water molecule is connected to other molecules are therefore of the greatest importance, and we seize with eagerness on the opportunity of studying them when they form a principal feature of a crystalline structure which the X-ray methods allow us to examine.

The main features of the crystalline structure of ordinary ice were revealed by X-ray methods of study about fifteen years ago. It is very interesting to bear them in mind when we review the old arguments and attempts at explanation. There is much to be said for a second reading of a detective story. The arrangement of the oxygen atoms can be simply described, though not, perhaps, quickly appreciated without the study of models such as those before you. (Plate V, A, B). Each oxygen lies at the centre of a regular tetrahedron of which the four corners are occupied by other oxygen atoms. The whole structure is hexagonal, differing in this respect from its companion structure, that of a diamond. In the latter the carbon atoms also follow the tetrahedral rule, but a difference in the general arrangement in the diamond puts the crystal into the cubic system.

The first attempt at a physical explanation of the so-called plasticity of ice was made by Tyndall and Huxley in 1857*. Faraday had drawn attention to the remarkable fact that two pieces of ice brought into contact with each other froze together. If pieces of ice floating in water were merely brought into contact, they became rigidly attached to each other. (Plate VI A). As Faraday said, the molecules of water change more readily from the liquid to the solid state when they are surrounded by a consider-

* *Phil Trans*, Vol cxlvii, p 327 1857

able mass that has already solidified. The thin sheet of water at and about the point of contact of two pieces of ice freezes and joins the two pieces rigidly, even though it would not freeze away from the ice. On this view freezing depends not only on temperature but on surrounding conditions also. In the same way water vapour within an ice crevice will settle on the ice surface and take its place there as part of the structure more readily than in the open where contact is possible with one ice surface only.

Tyndall and Huxley (loc. cit.) supposed that ice under tension would yield with the formation of cracks, these would, however, be healed by the deposition within them of vapour or liquid from neighbouring portions of the ice. Thus the mass of ice would be accommodated to the form of the bed in which it lay, and the glacier would 'flow'.

James Thomson* introduced a second principle. Accepting the fact that Faraday had demonstrated, he nevertheless rejected his mode of explanation. He ascribed the cementing of adjacent ice blocks to some pressure which forced the two together. Since water expanded on freezing, a pressure which tended to prevent the expansion would act against the freezing, in other words would lower the freezing point. Ice at 0°C should, therefore, liquefy under pressure. This also is easily observed. When a sharp point, such as that of a knife blade, is pressed on a block of ice it enters the ice, not because it cuts it or cracks it but because the ice melts before it. Thomson argued that when two pieces of ice are pressed together there is melting at the point of contact; the water so produced slips away on all sides and freezes again because it is no longer under pressure.

* *Proc. Roy. Soc.* Vol. x, p. 152. 1860.
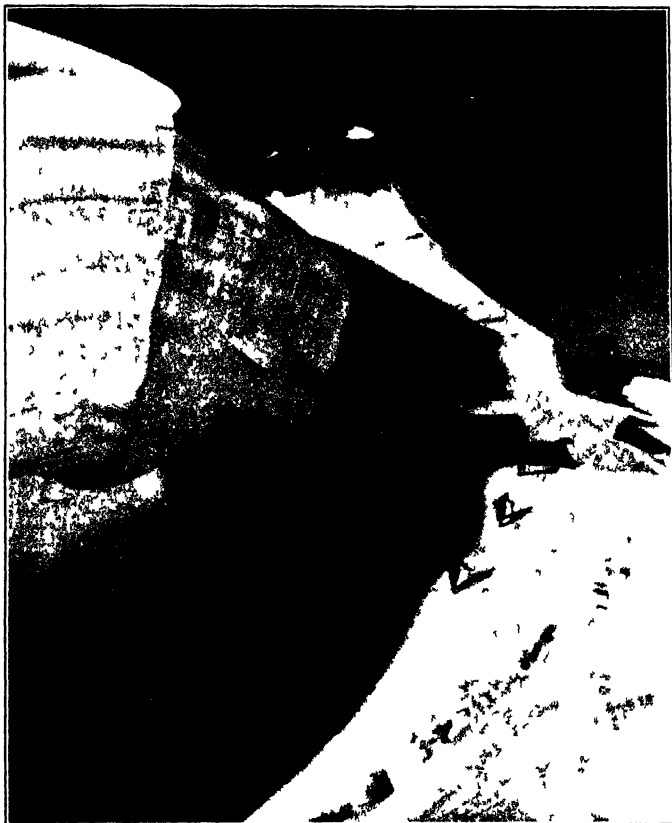
*Photograph by A Steiner, St Moritz*

PLATE II  Stratification of Firn Snow and Ice on Crast 'Aguzza
From Seligman's " Snow-Structure and Ski Fields "

Tyndall used to demonstrate this effect by crushing ice, which might originally be one piece or many pieces, in a mould under great pressure. The broken fragments froze together, all gaps being filled with the melted water which was restored to the solid condition. Tyndall's moulds are preserved and may once more be used; it is easy, for instance, to make a solid cup out of powdered ice under pressure of a few tons to the square inch.

The application to glacier flow is obvious. It was argued by Thomson and by Tyndall that the change of form of the moving glacier was brought about by melting in places where the pressure was great, and by the re-solidification of the escaping water.

Thomson's assertion that ice melted under pressure was, of course, perfectly sound. He was wrong, however, when he supposed that Faraday's regelation was to be explained on the pressure principle. He argued that when two pieces of ice were merely brought together without obvious pressure, there was a pressure due to capillary action and it was this that was the cause of Faraday's effect. To meet Thomson's objections Faraday* repeated his experiment under circumstances in which there could be no capillary action. He prepared a number of discs of ice, the upper surfaces being rounded to a convex form, and floated them convex face uppermost on the surface of water. Under such circumstances the blocks could touch under water only and not on their water lines (Plate VI A). When brought gently together they froze into a rigid mass which could be guided in movement over the water surface by pressure applied at one point only. If they were replaced by

*Proc. Roy. Soc. Vol. x. p. 440. 1860.

wooden discs, the latter, even if they could be made to touch on the water line and therefore attract each other by capillary action, could and did roll round each other and showed no trace of rigid connection. If, like the specially formed ice discs, they could only be made to touch below the waterline, they had no mutual attraction whatever. In another experiment Faraday attached lead sinkers to blocks of ice which he submerged in water. When the blocks‹ were brought into contact they froze together and remained in conjunction even when the sinkers were so placed as to exert a separating force (Fig. 1).



FIG. 1. A drawing from Faraday's notebook, showing some details of his arrangement for proving his principle of regelation. In the upper figure a vessel of ice-cold water is surrounded by an ice-pack. Two blocks of ice, anchored to lead sinkers can be made to lean over and touch each other, whereupon they freeze together and stay so against the pull of the cords attached to the sinkers. In the lower figure a block is shown enlarged.

Nevertheless, these principles, Faraday's and Thomson's, were unable to account satisfactorily for glacier motion; their effects were inadequate. Moreover they ceased to act at low temperatures, such as must certainly be found in some of the glaciers of the world. Faraday himself pointed out that two pieces of ice when cooled well below zero could not be made to unite by any pressure which he exerted.

Many years later the matter was again taken up, this time by McConnell,* a Cambridge graduate

* *Proc. Roy. Soc.* Vol. xlix. p. 323. 1891.

*Photographed by E. Meerkamper, Davos*

PLATE III   Summit Ridge of the Piz Roseg

From Seligman's 'Snow Structure and Ski Fields."

who was forced by ill-health to spend much of his time at Davos in Switzerland. Some preliminary experiments led him to conclude that the secret of the plasticity of ice could only be explained after a satisfactory examination of the properties of the single crystal. Most ice, particularly artificial ice and the conglomerate ice of the glacier, (Plate VIB) is a mass of separate crystals. McConnell cut single crystals of sufficient size from the ice frozen on the surface of a lake or in an open vessel. The one axis
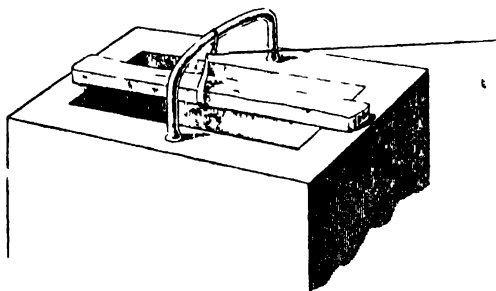


FIG 2   A drawing by McConnell of his arrangement for
observing the elasticity of an ice crystal.

of the crystal is always perpendicular to the surface of the open water on which the ice is formed. He examined the yielding of crystals under stress when (i) the specimen was a slab of ice a few inches long and an inch or so wide and deep, and the slab was laid on its supports (Fig. 2) with its upper surface disposed as it was when it was formed; (ii) when the same specimen was turned over through a right angle so that the 'lake surface' became vertical; (iii) when the specimen was a column which had been vertical in the original mass. The first two examples are illustrated by the models in Plate VII, A, B, C and the effects of loading are also shown. In the first case the specimen bent like an ordinary

bar under load. In the second the yield was of the same character but far smaller. In the third the deformation was of the peculiar character shown in McConnell's drawings (Figs. 3 and 4), and it was clear that there was movement on slip-planes. In the model (Plate VII D), the kind of movement is represented by the sliding of vertical glass plates held together by oil or other viscous medium. The arrangement in the figure had to be made by hand, there was no suspended weight. It is difficult, and in fact impossible, to make the model take up automatically the true mode of yielding, because
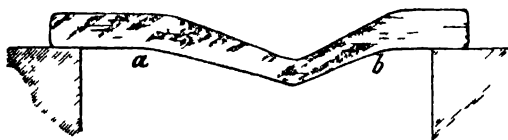


FIG 3. A drawing by McConnell showing the peculiar distortion, under a load, of a bar, the axis of which was vertical when it formed part of the surface ice on a lake

a few planes persist in taking all the slip. The forces between plate and plate are not so nicely balanced as in the ice. But the model illustrates the main point. In this third case McConnell's bar of ice yielded continuously and indefinitely, its movement was plastic. In the first two cases the yield was elastic, it reached a point and stayed there, the amount of yield being proportional to the weight.

A metallic mass, like a glacier, is in general an agglomeration of crystals in each of which there are slip-planes. The plastic deformation of the mass by bending or stretching without breaking is accomplished through slipping on various sets of planes. It is possible to prepare a single crystal of certain metals, for example copper, which will show the
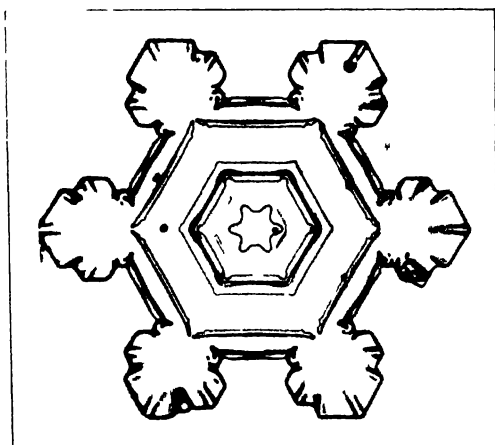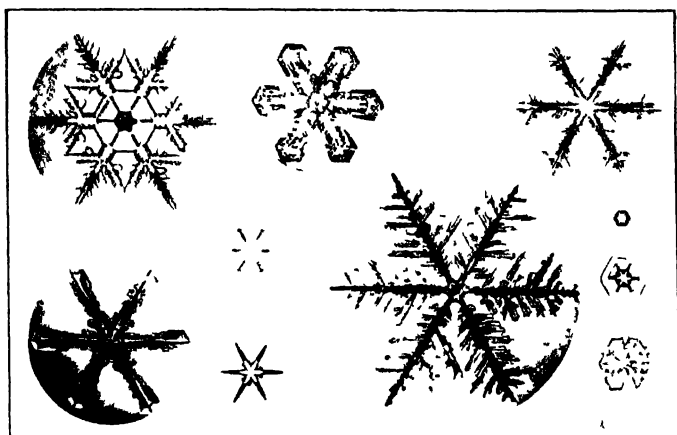
Photographs of snow-crystals by Nakaya.
J Fac Sc Hokkaido Imp Univ Ser a Vol. i

PLATE IV

nature of the movement very clearly. Such a crystal
may be an inch or two long and half an inch in
diameter. It can be bent between finger and thumb.
After one or two deformations, however, it becomes
highly resistant, and the fingers cannot bend it any
more. It is then found that the mass is no longer a
single crystal; crystals of all magnitude down to the
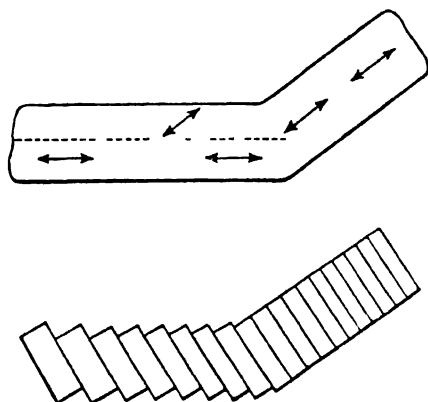sub-microscopic have been formed, and there is no



FIG 4. A drawing by McConnell showing his explanation of the
effect in Fig 3 as due to sliding on planes

longer a single set of slip-planes all parallel to one
another. This 'work-hardening' does not occur in
the case of ice.

The existence of slip-planes in the single crystal
suggests an explanation of glacier motion parallel
to that which is applicable in the case of plastic
metal. It is true that the regelation and the lowering
of freezing point by pressure must have an effect
which has no parallel in the case of the metal. On
the other hand ice is not hardened by work. These
are, however, differences of secondary importance.

In order to obtain a clearer conception of the
slip-plane action in ice let us turn to the model of

the structure obtained by the X-ray methods. It will be observed that the crystal may be looked on as a set of parallel sheets, joined by links which are at right angles to the sheets. Each sheet is puckered (Plate V B): its internal strength may be expected to be considerable, but its links to the sheets on either side of it are all parallel to one another and perpendicular to the general plane of the sheets. They may be expected to offer relatively small resistance to the displacement of one sheet in any direction in its own plane, even to such a displacement that links are broken and reformed between new pairs of atoms. The motion might be compared to that of the passage of a brush over a rough surface, during which the bristles continually yield and then spring back to make fresh contacts.

The sheets in the ice crystal are found to be parallel to the surface of the ice on the lake or vessel. In the model of the third case described above they are represented by the sheets of glass, which slide over one another. In the first two cases the sheets are represented by the wooden strips, which bend under load; the yield is, however, far less in the second case, where the strips are on edge, than when they lie flat as in the first case.

McConnell made an interesting observation on the behaviour of the ice in the first case. Where the water is being frozen on the lake surface, bubbles are formed which have the shape of thin vertical streaks. As the specimen of ice yields under load these bubbles remain vertical and therefore cease to be parallel to the axis of the crystal, which remains perpendicular to the surface at its different points. To represent this effect in the model (Plate VII, A, B), a vertical black line is drawn on the side of the bundle of laths. When the bundle is bent
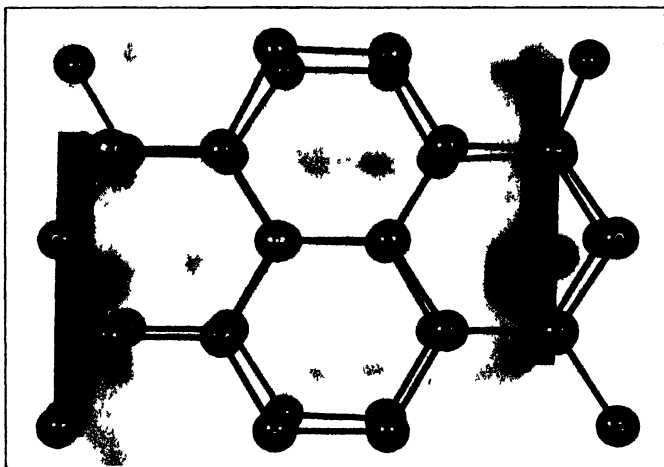
**A.** The ice model, viewed in a direction perpendicular to the sheets of B and therefore along the optic axis the hexagonal character is obvious
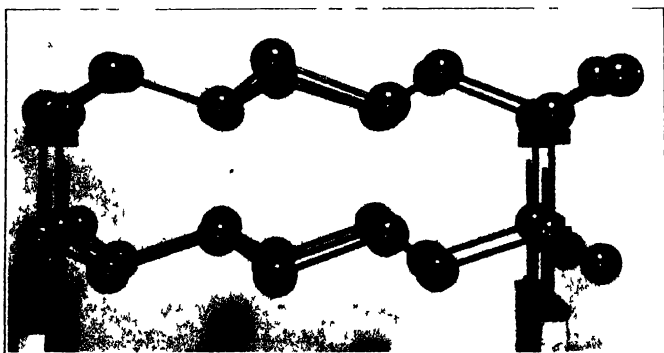


**B.** Another view in a direction at right angles to that of A the two sheets of the model are seen distinct from one another The crystal consists of repetitions and extensions of the sheets The distance from the centre of each of the oxygen atoms, represented by the black balls, to the centre of each of its three neighbours in its own sheet (A) is 2 74 Angstrom Units (100,000,000 Angstrom Units are equal to one centimetre) This is also its distance from an atom lying either immediately above it or below it in a neighbouring sheet The tetrahedral arrangement of any atom and its four neighbours is shown in Plate VI C, where the oxygen atoms are drawn large enough to touch each other The links connecting the atoms in the sheets are represented by rods the links between sheet and sheet are inserted in a few cases only The positions of the hydrogen atoms are not directly revealed by the X-ray analysis, and are under discussion

PLATE V.

under load, the line is broken into steps which lie on a vertical line: the steps are visible in the model but necessarily they are not to be seen in the ice. In this bending the laths slip on one another, again demonstrating the slip-plane mechanism.
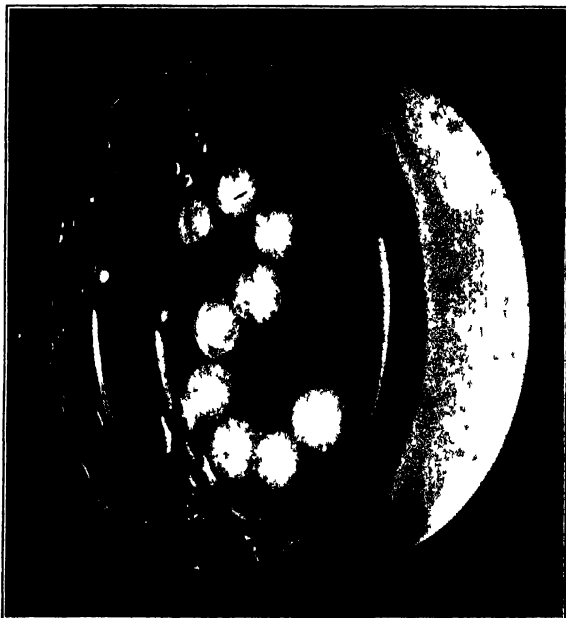
McConnell also observed that in the third case there was a certain slow recovery if the load was removed. Its total amount was such that the two sides of the V-shaped bar turned each of them through an angle of two or three degrees, the obtuse angle of the V being increased thereby. The links between sheet and sheet throughout the crystal must, on the average, have unbent by this amount during the recovery which he observed.

The ice crystal, unlike the metal, is not hardened by cold working. The sheets as they slide past one another must remain undistorted and unbroken, so that the condition of the crystal is not altered. In the metal there is atomic disarrangement and the crystalline conditions become more complicated.

Faraday's regelation and Thomson's liquefaction by pressure are both consequences of the open structure of ice, which again is a consequence of the special character of the bonds that bind its atoms and molecules together. Of late years there have been considerable advances in our knowledge of atomic bonds in general. For our present purpose it is sufficient to observe that in some cases an atom attaches other atoms to itself only at special points on its surface; it may be more correct to say that the points at which atoms are attached are distributed over the surface of the atom according to rule. Thus the carbon atom attaches to itself four other atoms at points which lie on the surface at the four corners of a regular tetrahedron. In
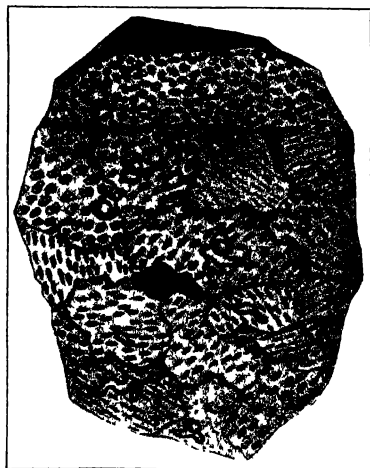
other cases there is little or no directional arrange-
ment of the attracting forces, as for example, in
liquid mercury, in fact the intrinsic property of
any liquid is due to the absence or the masking by
stronger forces of any directional forces that would
build up a structure. One might imagine extreme
cases. In one of them, there would be no forces
except those of the first kind; the atoms would have
no mutual attraction except when so presented that
they locked at the definite locking points which each
possessed. An assemblage would then be either a
solid or a crowd of individuals with no mutual
attraction. In the latter case the assemblage, if
sufficiently agitated by its heat content, would take
the form of a gas. Under those circumstances, one
might suppose, occasional lockings would take
place. If the heat were diminished the lockings
would become more frequent and more perman-
ent, and solidification would occur here and there.
Each solid assemblage would be a crystal unless the
process was so hurried that the atoms or molecules
did not settle truly into a regular crystalline
arrangement. They would settle, preferentially, on
aggregates already formed, rather than on stranger
bodies, because they could fit into places naturally
ready for them. Thus the crystals would grow out
of the vapour, without the formation of a liquid
intermediate stage. In the extreme case of the
opposite kind, one might imagine atoms or mole-
cules which had merely a general attraction for
others of their own kind. When the agitation due to
heat was insufficient to keep them separate, they
would draw together as a liquid which would pos-
sess a viscosity depending on the ease with which
individual molecules could move about in the
assemblage. With further cooling the viscosity

B   Drawing of a section of glacier ice by Agassiz showing its conglomerate character    The 'ice flowers' (Plate VIII F, G) have different orientations in the different fragments

C   See description under Plate VB





A.  Faraday's experiment   the floating blocks are frozen together on being brought gently into contact, whereupon the set forming a letter S can be moved as one whole

PLATE VI

might increase continuously or by sudden steps, to such an extent that the assemblage could be called a solid.

There is an interesting passage in Newton's *Opticks* in which two such different modes of atomic attraction are vaguely alluded to. Newton

The Parts of all homogeneal hard Bodies which fully touch one another, ftick together very ftrongly. And for explaining how this may be, fome have invented hooked Atoms, which is begging the Queftion; and others tell us that Bodies are glued together by reft, that is, by an occult Quality, or rather by nothing; and others, that they ftick together by confpiring Motions, that is, by relative reft amongft themfelves. I had rather infer from their Cohefion, that their Particles attract one another by fome force, which in immediate Contact is exceeding ftrong, at fmall diftances performs the chymical Operations above mention'd, and reaches not far from the Particles with any fenfible Effect. ‑ ‑ ‑ ‑

‑ ‑ ‑ ‑ There are therefore Agents in Nature able to make the Particles of Bodies ftick together by very ftrong Attractions. And it is the Bufinefs of experimental Philofophy to find them out.

FIG. 5. Newton's *Opticks*, 3rd Edn., pp. 364 and 369.

very properly pointed out the weakness of the suggestions made in his day, but the terms used in those suggestions might now be used with more significance (Fig. 5). We may not inaptly distinguish between atoms provided with hooks and others that are not.

Natural substances lie between the two extremes, ice being nearer to the former than is usual. Crystals form directly from the vapour. On the other hand

molecules of water vapour are continuously being torn away by heat or other cause from the crystal and adding themselves to those in the surrounding space. The ice 'sublimes'. There must be a relation between the density of the vapour, the temperature, and the surface conditions, which determines the ease or difficulty with which a molecule settles in its place or conversely breaks away. In an ice crevice the chances of settling are greater than in the open beside an ice face because, beginning at the fine edge of the crevice, the molecules have a double anchorage, and there is throughout a greater chance of a retention of the molecule when the ice surface is concave. The effect is exactly analogous to that of the dependence of vapour pressure on the form of the liquid with which the vapour is in contact, as indeed Faraday pointed out. In the crevice the balance between sublimation and deposit is not the same as in the air outside, the latter is increased at the expense of the former. In other words, the vapour pressure is lowered, or the freezing point is raised.

When water freezes and the open ice structure is formed, there is an expansion which can be hindered to some extent by pressure. Thomson's principle expresses the fact that pressure crushes the structure altogether, breaks certain of its bonds and so liquefies it.

Under pressures of ordinary magnitude, water is at its densest at a temperature of 4°C. At lower temperatures than this it becomes less and less dense as the freezing point is approached. Clearly some of the molecules are forming associations which take up more room than the separate individuals, because the linkages force the atoms to take up a definite arrangement in space. It is not neces-
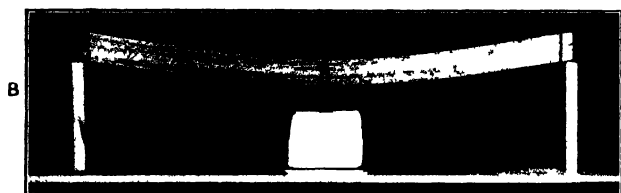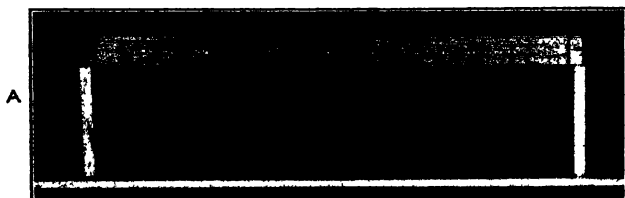
**A** Model made of laths showing **B** the bending under a weight as in McConnell's first experiment The vertical black line towards the left of the bundle remains apparently a continuous line after the bending it is actually zig-zag The black line at the centre is merely a shadow of the wire supporting the weight
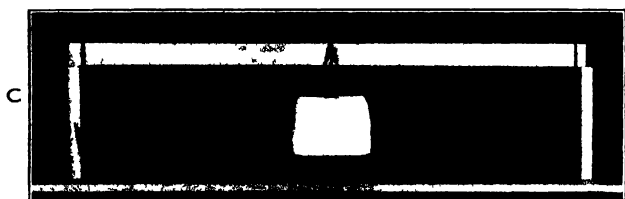


When laths are placed on edge the bending is very slight



**D** A model illustrating the effect in Fig 3 A series of glass plates are held together by films of glycerine and, for purposes of photography, by a bent strip they imitate the action of slip-planes

PLATE VII

sary that the associations should be permanent; probably they are constantly made and remade.*

Let us now consider the snowflake. (Plate IV). The two striking features which require explanation are the immense variety of the hexagonal forms and the symmetry of each flake. The hexagonal character is clearly related to the hexagonal form of the ice crystal, but it has always been difficult to explain why the six arms in each flake are so like one another and at the same time differ so widely from the arms of every other flake. It would appear probable that a sufficient explanation can be given in the following way. The whole matter is related to the same effects of sublimation and deposition as are concerned in regelation. When the hexagonal construction grows outwards from some nucleus in an atmosphere of water vapour, or in a liquid in which water is dissolved, the character of the growth is affected by the conditions under which it takes place. If there is much vapour or a low temperature the growth is rapid, the flake grows quickly outwards from the centre, always reaching out to the richer surroundings from which the molecules have not yet been abstracted and deposited. Accretions take place at the ends of the radial arms, or the subsidiary arms which spring from them at an angle of 60°. But if the flake is transferred to a place where the vapour density is less or the temperature is higher, the reverse procedure sets in. Molecules are more easily detached from the ends of the arms than from the central regions of the flake, and deposited less readily. Thus the flake now consolidates in its central portions and its more angular extensions may even be shortened or rounded off. The flake is quite small,

* Bernal and Fowler, *J. Chem. Phys.* Vol. I. p. 515. 1933.

not often more than a few millimetres in diameter, so that the conditions round the flake are the same at all points. If one arm is growing at its extremities, so are the others; if the recent spaces nearer to the centre are being filled in, the same is happening to all six arms. But of course the flake may go through a great variety of local conditions on its descent to the earth, and the arms will assume a corresponding variety of forms, while the six arms on any one flake are all alike.

On this explanation the rich variety of snow forms is based on the existence of two opposing processes, one of them a rapid extension of radial arms, leading to a feathery form, the other a consolidation in which the centres of the feathery extensions are filled up. The first process involves relatively more accretion and less sublimation than the other, so that they differ in degree more than in kind. The form of each snowflake tells the story of the variations in the conditions of the atmosphere through which it has passed on its way to the earth. A still slower growth of ice crystals may be made by keeping ice at a temperature just below freezing point in one arm of an exhausted vessel, and allowing it to sublime on to a metal plate placed in another arm of the vessel kept at a considerably lower temperature. The crystals take the form of solid hexagonal pyramids or frustra. (Plate VIII E.)

The difference may be demonstrated experimentally. In Plate VIII the feathery forms of A were obtained by rapid cooling of a mixture of glycerine or alcohol and water, in which the water was relatively a large constituent. The more solid looking hexagons of B were obtained by cooling more slowly a mixture containing less water. When the hexagons obtained in this way are allowed to
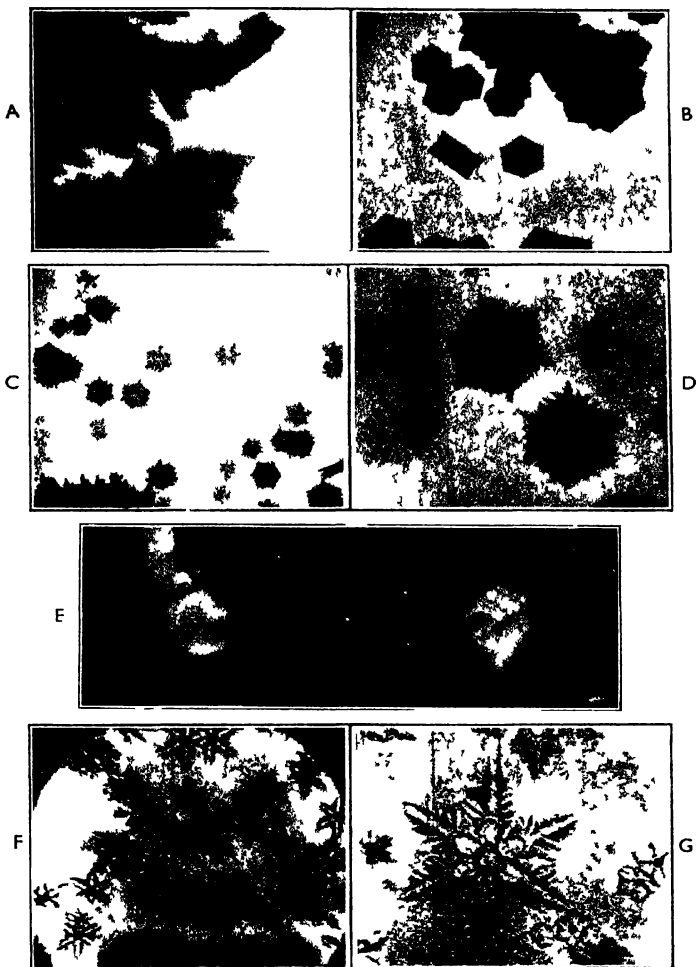
A Ice crystals formed in a mixture of water (85° ) and glycerine (15 %)
it a low temperature  Observe the feathery character  B Ice crystals
formed in water (66° ) and glycerine (34  ) at a slower rate than in
A  Observe the hexagonal tabular character  C The tables of B being
allowed to melt slowly, reveal a structure resembling a true snow
crystal  The tables are in various stages of disappearance  D An en-
larged picture of some of the crystals of C   The slight haziness of
outline in C and D is a consequence of difficulties of focussing
through the liquid  E  Hexagonal pyramids of ice formed by sublima-
tion  F  Ice flowers formed in a block of ice under radiation  Observe
     the similar orientation of all the flowers  G  A single ice flower

PLATE VIII

melt or sublime, it is found that there is an under-lying and more open structure, which resembles quite closely the natural snowflake in some of its forms. (c and d). These processes can be watched upon the screen.

Tyndall used to show a beautiful experiment by which he demonstrated the ice structure. It was known as the experiment of the " Ice Flowers." A beam of light from an arc lantern was passed through an ice block. At various points in the block
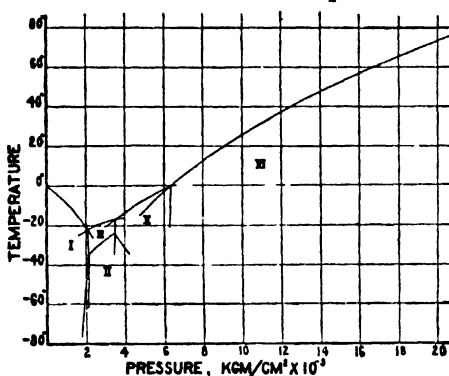


FIG 6   Ice phase diagram due to Bridgman

where there were impurities or irregularities suffic-ient heat was intercepted to cause melting. The cavities thus formed were hexagonal in form be-cause the structure came apart in the reverse order to that in which it had been put together. They were, so to speak, negative snowflakes, less fine in outline than true flakes, but nevertheless shaped in the same fashion. By an optical arrangement the cavities could be focussed on a screen.

A film which has been prepared in the Institu-tion Laboratory shows the phenomenon very well. An extract is given in Plate VIII, F, G. In this case the whole of the block under investigation is one

crystal, and the six pointed figures are therefore similarly oriented throughout. As the beam from the lantern warms the ice the 'flowers' grow and take on richer forms. If a water-cell is placed in the path of the beam it absorbs the heat and the growth of the flowers is arrested. A single flower is shown in Plate VIII G.

Bridgman has shown* that under great pressures ice changes its structure. Several different forms can be recognized, and in fact a complete phase diagram can be drawn. (Fig. 6.) Two of these new forms, Ice II and Ice III, have been examined by



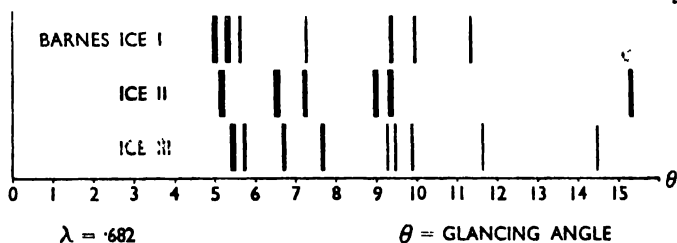$\lambda = \cdot 682$       $\theta$ = GLANCING ANGLE

Fig. 7. X-ray spectra of Ice I, Ice II and Ice III, the latter two drawn from the observations of McFarlan. The principal lines only are shown.

the X-ray method by McFarlan†, who has suggested structures in accordance with his observations. The principal lines of the X-ray spectra of Ice I (ordinary ice), Ice II and Ice III are shown in Fig 7 : the latter set out in accordance with McFarlan's results. It is clear that the three forms of ice are crystalline and that their structures are very different. McFarlan's Ice II retains the puckered sheets practically unchanged, but the sheets are forced together by a bending of the bonds between them. The distance from oxygen centre to oxygen centre—the length of the bond—remains

---

* *Proc. Amer. Acad.*, Vol. xlvii, p. 441. 1912.
† *J. Chem. Phys.*, Vol. iv, pp. 60, 253. 1936.
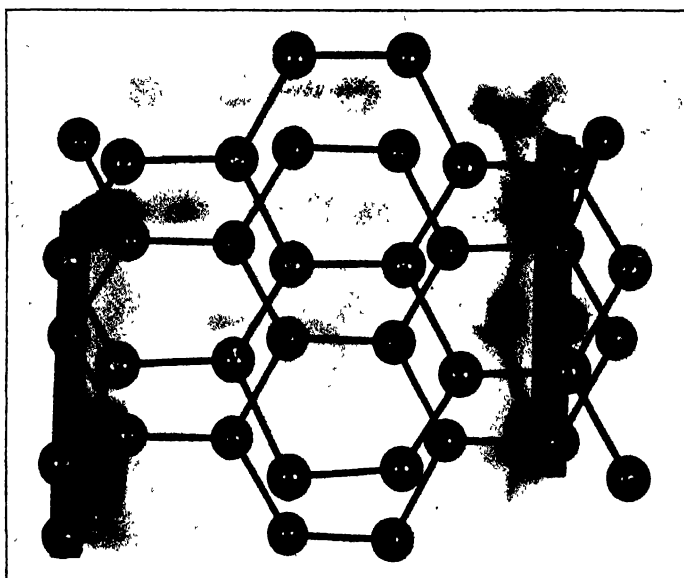
A
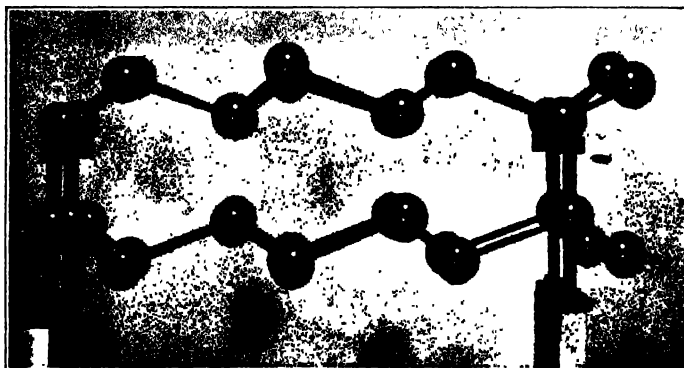


B

**A and B.** The model of Ice II. Comparison with the Plates V A and B will show that the sheets are unaltered, but are brought nearer to each other by pressure. Models made in accordance with the results of McFarlan.

PLATE IX.

unchanged. The density of the crystal is about 1.21.
Ice III is also crushed together, but in this case it
is the sheet that gives way, the distance from sheet
to sheet remaining almost unaltered. As before, the
distance from centre to centre of neighbouring
oxygens has still the same value.

The new structures are best studied with the aid
of models. Photographs of the Ice II model are
shown in Plate IX A and B: they are necessarily
more difficult to visualize than the model itself.



Density—0.90        Density=1.21        Density=1.105
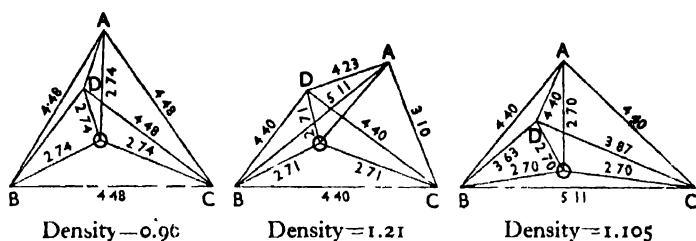
FIG. 8. The oxygen tetrahedra in the three forms of ice, drawn
in accordance with the results of McFarlan. In each case O, lying
at the centre of the tetrahedron is equidistant from the four
corners A, B, C, D.

Ice III is worse than Ice II. Perhaps the simplest
way of exhibiting the main features of the differ-
ences in structure is to represent the tetrahedra
which in each case surround the oxygen atom.
This is done in Fig. 8. The dimensions are in accord
with the observations of McFarlan.

I am much indebted to Mr. Green and his
assistant Mr. Bridger for their care and skill in the
preparation of the experimental illustrations of
this lecture.

[W.H.B.]

# Friday, May 6, 1938

SIR ROBERT ROBERTSON, K.B.E., M.A., F.R.S.
*Treasurer and Vice-President, in the Chair*

H. W. MELVILLE, D.Sc., Ph.D.
*Fellow of Trinity College, Cambridge*

## THE NATURE OF CHEMICAL REACTIVITY

IN this discourse it is my intention to discuss the present ideas of the nature and mechanism of chemical reactions. The subject of reactivity is, of course, as old as chemistry itself and, if we may judge from the volume of current literature, partly dominates the science. Consequently only one or two aspects of the problem can be dealt with in the time available. Fortunately one restriction arises almost automatically. Reactions are usually simplest in the gas phase because of the fact that molecules are normally separated from each other by distances much greater (a thousand times at atmospheric pressure) than their own dimensions. We shall therefore confine the discussion to gas reactions and for that purpose the traditional test-tube and reagent bottle of the chemist must be replaced by apparatus slightly more complicated, to be described below, in order to demonstrate a few of the fundamental phenomena involved.

Perhaps one of the most striking characteristics of chemical reactivity is the enormous variation in the rate of reactions met with in practice. For example, if a piece of filter paper is dipped into a solution of phosphorus in carbon disulphide and the latter allowed to evaporate, the finely divided phosphorus immediately catches fire. On the other

hand finely divided sulphur—flowers of sulphur—does not oxidise appreciably when left in contact with the atmosphere for days and even years. There is no doubt that the molecules of oxygen collide with the solid sulphur and phosphorus at the same rate, and consequently collisions with sulphur must be very much less efficient than those with phosphorus. From this simple experiment then, the first general principle to be established is that the rate at which molecules collide with each other does not wholly determine the velocity of chemical reactions. The question arises: What is the reason for this chemical inertia? Returning again to the oxidation of sulphur, it is a common observation that if the temperature of the system be raised to some 100° C, oxidation will occur at a measurable velocity. That is by raising the temperature of a system, chemical inertia may be overcome. It might be argued that the increase in velocity is simply due to an increase in translational velocity of the gas molecules, but the increase of rate of nearly all reactions with temperature is far greater than the increase in molecular velocity. Those molecules which react must therefore be in some exceptional state, and in order to explain this observation it was suggested many years ago by Arrhenius that only molecules possessing a large amount of energy could so react. Now according to the kinetic theory of gases molecular energies are not all equal but in fact are grouped around a mean value, which is determined by the temperature of the gas. Arrhenius considered that an equilibrium was set up between the normal molecules N and activated molecules A; thus N + energy = A. From the rate of increase in velocity this so-called energy of activation could be deter-

mined. But the significant fact was that the kinetic theory indicated precisely the same increase in the concentration of highly energetic molecules, establishing thereby a correlation between the chemical and physical phenomena which is of the utmost importance for the interpretation of reactivity.

Highly energetic or activated molecules thus play a predominant part in chemical reactions, and the next problem is to define more precisely what is meant by the term and also how such molecules may most conveniently be produced. (For the purposes of discussion the term molecules will be taken to include free atoms.) Activated molecules may be divided into four classes corresponding to the different kinds of molecular energy now recognised: (*a*) molecules in which one or more of the electrons have been excited to a higher energy level; molecules possessing (*b*) translational, (*c*) vibrational, or (*d*) rotational energy in excess of the mean value. In the latter two classes the energy may be so large that the molecules are dissociated to free atoms or radicals.

In order to demonstrate how such activated molecules can be produced use may be made of sodium vapour, since the presence of activated atoms is readily seen owing to their emission of the familiar yellow D-light. The energy required for excitation is comparatively large because it is necessary to heat the vapour to about 2000° C before appreciable emission of D-light occurs, as is easily demonstrated by injecting a sodium salt into the air stream feeding a bunsen burner. This then is the first method of producing activated molecules —merely by raising the temperature of the whole system. 2000° C is, however, rather a high tempera-

ture to use for chemical experiments, especially if the gases have to be contained in gas-tight reaction vessels. In such circumstances, where comparatively high excitation energies are involved, it therefore becomes necessary to employ some other method of excitation so that the reaction may be carried out at a lower temperature. A convenient, though from the chemical point of view rather drastic method, is to pass an electric discharge through sodium vapour when collisions between normal atoms and ions and electrons excite the
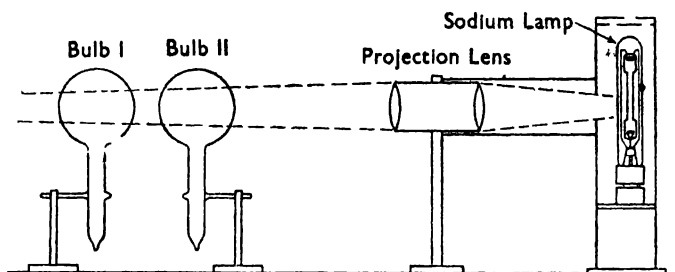


FIG. 1. Sodium fluorescence experiments

former to emit D-light as is shown best by the sodium lamp in fig. 1. Here the lamp consists of an inner discharge tube fitted with two filaments and filled with neon to a pressure of a few millimetres, and metallic sodium. The discharge first passes through the neon, but as the tube warms up the sodium D-light is emitted. To maintain the temperature of the inner tube at the optimum value, it is enclosed in an evacuated envelope. The trouble about the electric discharge is that it excites any other atoms or molecules present in the system and thus complicates the determination of reaction mechanism. It will be shown below how this difficulty can be partially overcome. The compli-

cation can also be disposed of by producing excited atoms photochemically by projecting a beam of light from the sodium lamp, as shown in fig. 1, into a bulb containing sodium metal heated to about 200° C by bunsen burners, when the vapour fluoresces very brightly. (Care must be taken to heat the bulb uniformly so that there are no cold patches to prevent the vapour pressure rising to the requisite value. At the start of the experiment the metallic sodium is best distilled into the side tube. The bulb is then heated, followed by the side tube. Too strong heating causes complete absorption and confines fluorescence to a layer within a few millimetres of the surface of the bulb facing the sodium lamp.) This then is the simplest type of photochemical reaction in which the valency electron of an atom is merely raised to a higher energy level. That such excited atoms can be used to induce chemical reaction may be demonstrated in the following way. A second similar bulb is placed in front of the first, and, keeping the latter at a constant temperature, this second bulb is gradually heated, when the fluor-escence passes from bulb 1 to bulb 2. Eventually the fluorescence in bulb 1 is completely extinguished. This is shown most strikingly on removing bulb 2 when bulb 1 immediately fluoresces as before. Next bulb 3, containing in addition to sodium, hydrogen at a pressure of 10 mm., replaces bulb 2. When this bulb is heated the fluorescence in bulb 1 disappears, but only reappears in bulb 3 very feebly, as is again shown by removing bulb 3. The hydrogen thus decreases or quenches the fluor-escence of the sodium vapour. A simple numerical calculation reveals immediately the reason for this behaviour. The lifetime of the excited atom is only

$10^{-7}$ second. In the absence of any complications, an atom would radiate its energy after this period elapses. But in the presence of hydrogen at 10 mm. the excited atom collides with a hydrogen molecule within $10^{-8}$ second and thus the probability of the latter process is ten times that of the former. Radiation of D-light is therefore largely prevented by this chemical interaction of Na and $H_2$ which produces sodium hydride and a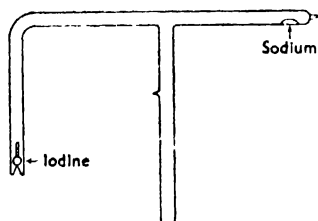 hydrogen atom thus:—$Na + H_2 = NaH + H$. It is normally impossible to produce hydrogen atoms photochemically because hydrogen only absorbs at a wavelength of about 1000A., which is much too short to be usefully employed in photochemical experiments owing to the fact that atmospheric oxygen exhibits absorption at about 1900 A. This type of reaction is termed a sensitised decomposition of hydrogen.



FIG. 2. Sodium-iodine chemiluminescence.

In some chemical reactions the reverse type of reaction occurs, that is the emission of light as a result of chemical combination or chemiluminescence. The majority of such reactions are so complicated that their precise mechanism is unknown, but in the experiment to be described below, namely the vapour reaction between sodium and iodine, matters are sufficiently simple that the mechanism has been elucidated. The apparatus, constructed of Pyrex, is shown in fig. 2. First a small evacuated bulb containing iodine is placed in the left limb of the apparatus. Sodium is distilled in at the other end, the whole evacuated and sealed

off. Before starting the experiment the iodine bulb is broken on the re-entrant tip of the side tube. The sodium is then heated strongly and the iodine gently. At the zone—1-3 cm. long—where the vapours interdiffuse, a bright emission of D-light is obtained simultaneously, accompanied by the deposition of sodium iodide. First a sodium atom attacks an iodine molecule, $Na + I_2 \rightarrow NaI + I$; next the iodine atom attacks a sodium molecule, $I + Na_2 \rightarrow NaI + Na$. This latter sodium iodide molecule possesses a large amount of energy which may be transferred in *one* collision to a normal sodium atom, thus exciting the atom and so causing it to emit D-light. This is an alternative way in which the energy liberated in a chemical reaction is removed from the products to that which occurs normally where the energy is dissipated throughout the system, with a consequent rise in the temperature of the reacting gases.

In the classification of active molecules there was included free atoms which play a large part in many gaseous reactions. The most convenient atom with which to work is hydrogen, for it is possible to demonstrate visibly some of its reactions and also some of the peculiar properties of free atoms. In order to do this, large atom concentrations are necessary, and for this purpose the electrical method of excitation has no equal. It is a very fortunate circumstance, which will be explained later, that the lifetime of atoms is very much larger than that of activated molecules or even ions. The periods concerned are about 1/10 second compared with $10^{-6}$ second. Hence if hydrogen at a pressure of about $10^{-1}$ mm. is passed through a suitable discharge tube at a linear velocity of flow of about 10 cm. per sec.,

atoms, activated molecules and ions will come out
of the discharge, but owing to the comparatively
low velocity of flow only the atoms will escape an
appreciable distance from the discharge.

The apparatus for the experiments is shown in
fig. 3. Hydrogen from a cylinder passes through



FIG. 3. Atomic Hydrogen Apparatus.
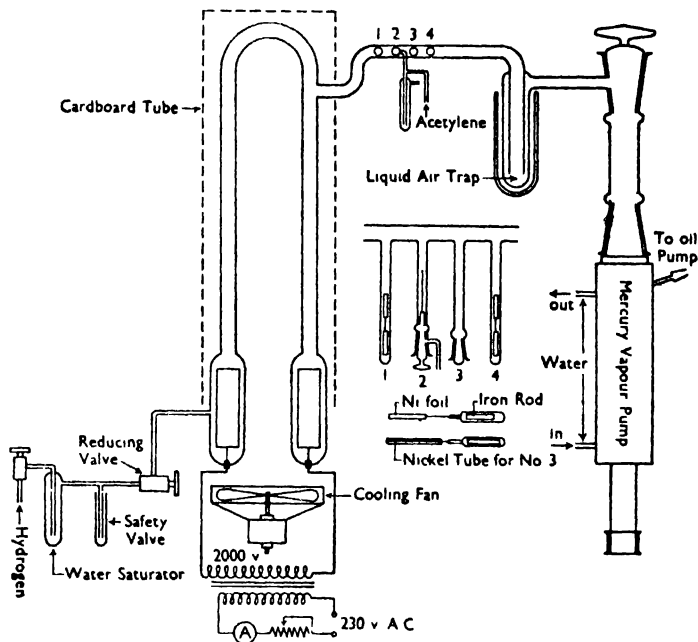
a gas wash bottle to saturate it with water vapour,
past a mercury safety valve and through an adjust-
able brass-steel needle valve which reduces the
pressure from atmospheric to about 0.1 mm. in the
discharge tube (2 cm. in diameter) constructed
wholly of Pyrex and containing aluminium elec-
trodes. The tube is excited from a transformer
giving at least 2000 v. and a current of 300-400 ma.

Cooling is effected by means of a small electric fan. The discharge tube is enclosed in a cardboard tube to prevent glare. The reaction tube is also a 2 cm. tube with a bend in it so that the light from the discharge does not enter the reaction system. The continuation of the reaction tube leads into a liquid air trap and then through a tap of 2 cm. bore to a mercury diffusion pump, having a speed of at least 10 litres per second. An oil pump of adequate capacity is necessary to cope with the output from the mercury pump. With a three-stage diffusion pump, a Hyvac or similar oil pump is suitable. Attached to the reaction tube there are four side tubes 0.5 cm. in diameter. The first contains a piece of nickel foil 0.1 mm. thick, spot-welded to a platinum wire fused into a sealed glass tube containing a length of iron rod, thus permitting the nickel to be slid into the stream of atomic hydrogen. The fourth tube is of similar construction. The second tube consists of a tap to which there is attached, by means of vacuum wax, a capillary tube a few tenths of a millimetre in diameter and about 5 cm. long as shown in the figure. By opening the tap the barrel of it may be filled with acetylene from a cylinder, thus causing a slow stream of acetylene to mix with atomic hydrogen. The third tube is fitted with a ground joint to admit a nickel tube 1 mm. in diameter, likewise attached to a glass tube enclosing an iron rod. Sodium is pressed into the nickel tube, the whole being immediately inserted into the apparatus which is then exhausted. When the pressure is adjusted correctly the colour of the discharge should be a fiery purple and a pocket spectroscope should only show the Balmer series without much trace of the molecular spectrum.

That hydrogen atoms do emerge from the discharge is easily shown by inserting the first piece of nickel foil into the stream, when it is heated white-hot within a few seconds. If the first piece is withdrawn and the foil in the fourth tube inserted, it is heated bright red showing that the concentration of atoms is not qute so high some 15-20 cm. along the tube. The atoms thus combine with each other either in the gas or on the walls as they traverse the reaction tube. Since nickel is a good catalyst for the dissociation of hydrogen molecules, it must therefore be a good catalyst for the association of atoms and hence the majority of the atoms combine on the foil, which is heated white-hot owing to the fact that no less than 100,000 calories per mole of hydrogen are evolved on combination. At a pressure of 0.1 mm. in the reaction system a hydrogen atom will collide with another atom in $10^{-6}$ second. It has, however, been demonstrated that hydrogen atoms have a much longer lifetime, namely $10^{-1}$ second. It is evident, therefore, that free atoms do not combine to form a molecule every time they collide. In point of fact, a hydrogen molecule can only be formed when a third body, such as a wall or another molecule or atom, is present to remove partly the energy of combination and so stabilise the molecule. If this energy is not removed the molecule merely dissociates to atoms once more. This type of three body collision may be visibly demonstrated in the following manner. The nickel tube containing the sodium is slid into the atomic stream and in a few seconds the familiar yellow glow appears in the *gas phase* around the end of the tube. What happens is that some of the hydrogen atoms combine on the surface of the tube,

thereby heating and vaporizing the sodium. The sodium atoms then function as third bodies in the gas phase for the removal of the energy from the quasi-hydrogen molecules. But in removing this energy the sodium atom is simultaneously excited, for only 50,000 cals. are necessary for excitation and 100,000 are available from the hydrogen molecule. In due course the atom emits D-radiation.

With the same apparatus CH and $C_2$ molecules may be produced from almost any hydrocarbon. Acetylene is one of the most suitable. When a dose of acetylene is admitted to the barrel of the tap in the second tube a bright green glow due to $C_2$ molecules appears in the reaction vessel, a fainter blue glow near the discharge tube is due to CH molecules. If the nickel foil in the first tube is inserted into the atomic stream, the glows are immediately extinguished, showing that hydrogen atoms are responsible. The precise manner in which the acetylene molecule is dissociated and the fragments subsequently excited is not yet known. This method of introducing gas into a stream of atomic hydrogen thus eliminates the difficulty mentioned previously about activating only one component of a mixture. Matters are so arranged that no appreciable amount of acetylene diffuses back against the stream of atomic hydrogen into the discharge. The technique is, of course, only applicable to flow systems.

One of the simplest reactions of atomic hydrogen is that with a deuteride, for example $ND_3$, thus $H + ND_3 = ND_2 H + D$. Unfortunately it is not an easy matter to demonstrate this reaction and so use must be made of models to show an interesting point regarding chemical reactivity. If the reaction, for simplicity, be regarded as $H +$

$XD = XH + D$, where $X = ND_2$ and it is assumed that exchange occurs by the hydrogen atom approaching the XD molecule from the side remote from the D atom, then three parameters characterise the state of the system. These are the distance between X and H and that between D and X. In a simple model such as fig. 4A, therefore, the occurrence of reaction
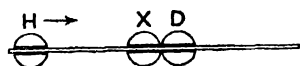


FIG. 4A.
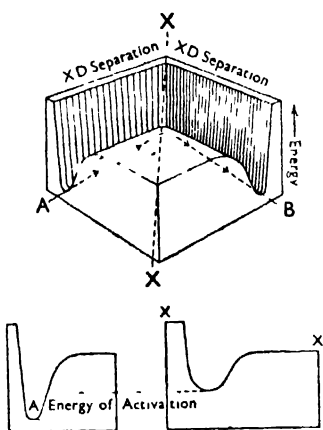




FIG. 4B.

occurrence of reaction may be represented by the movement of an H atom towards XD, the formation of an XH molecule and the removal of a D atom. Like most chemical reactions, this process possesses considerable inertia and therefore, the hydrogen atom must acquire a certain amount of energy before it succeeds in knocking the D atom out of the molecule. The third parameter of the system is therefore its energy, and in order to represent this variable it is necessary to use a three dimensional model. Accordingly, the two distances are plotted as rectangular co-ordinates and the potential energy of the system perpendicular to both. Now to each configuration of the system of atoms, there is only one value of the energy and consequently an energy surface will be obtained as shown in fig. 4B. Points at A and B at the extreme edges of the model at the bottom of the valleys represent the initial and final states of the system.

The reaction may thus be represented by a *point* moving over the surface from the initial to the final state. It is evident from the model that there is an indefinite number of reaction paths: in practice the easiest one is taken. If the model is divided into two along the dotted line XX, it will be seen that the difference in height between the positions of minimum energy represents the energy the hydrogen atom has to acquire before exchange will occur. The model can now be somewhat simpli-
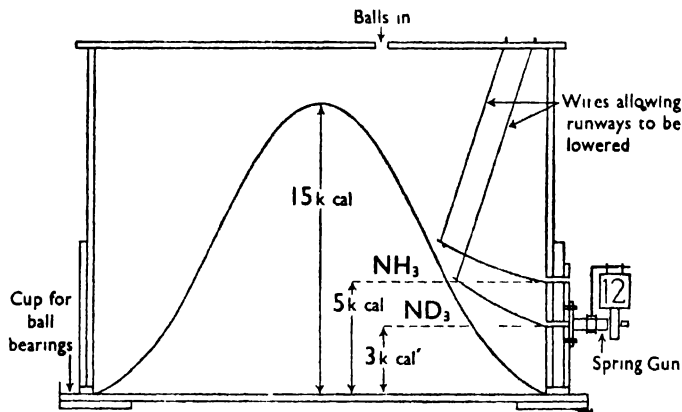


FIG. 4 C

fied by flattening out the reaction path so that the energy of the system is plotted vertically and the distance along the path horizontally. This is shown in fig. 4C. The reaction point is a steel ball. A spring gun attached to the right side of the apparatus can impart to the ball energies of 10, 12 and 15 k.cal. The height of the potential barrier, as it is called, is 15 k.cal. and hence if the ball is given an energy of 10 k.cal. it fails to surmount the barrier. On the other hand, when given an energy of 15 k.cal. it just succeeds in getting over the barrier.

When the velocity of the two exchange reactions $H + ND_3 = ND_2H + D$ and $D + NH_3 = NH_2D + H$ are compared it is found that the energy of activation for the former is about 2000 k.cal. greater than that for the latter, thus causing the reaction of the deutero-ammonia to be the slower of the two. To explain this observation it is necessary to invoke the principle of uncertainty, which plays a fundamental part in modern physical theory. One form in which the principle may be stated is that if the position of a particle be specified exactly, then its velocity is not known exactly and vice versa. One consequence is that a system of atoms, such an ammonia molecule always possesses a certain amount of vibrational energy no matter to what temperature the molecules may be cooled— the so-called residual or zero point energy. This energy amounts to no less than 3000 cals. for $ND_3$ and is somewhat higher for $NH_3$—5000 cals.— because the hydrogen atoms being lighter, vibrate rather more quickly than the deuterium atoms. At the temperature at which exchange is normally observed the energy possessed by the ammonia molecules is practically wholly residual energy. The system of atoms thus starts off with an initial energy of 3000 cals. for $ND_3$ in addition to that supplied by the H atom. The energy which the H atom has to acquire is thus reduced from 15000 to 15000 — 3000 = 12000 cals.* This is easily demonstrated by raising the gun a height of 3000 cals. when a particle of energy of 12000 cals. surmounts the barrier. For the reaction of deuterium atom with ammonia the gun is raised to a height

* In the exact treatment of the effect of residual energy account must be taken of residual energy at the top of the barrier which *increases* the activation energy. In general the increase is less than the decrease in the initial state of the system.

of 5000 cals. when a particle of energy of 10000 cals. then surmounts the barrier. The effect of residual energy is therefore to reduce the energy of activation of reactions, thus causing hydrogen containing compounds to react more quickly than those containing deuterium.

Mention has already been made of the fact that the lifetime of atoms is so short that except in a flow system, such as was used for the hydrogen atom experiments, this period cannot be directly
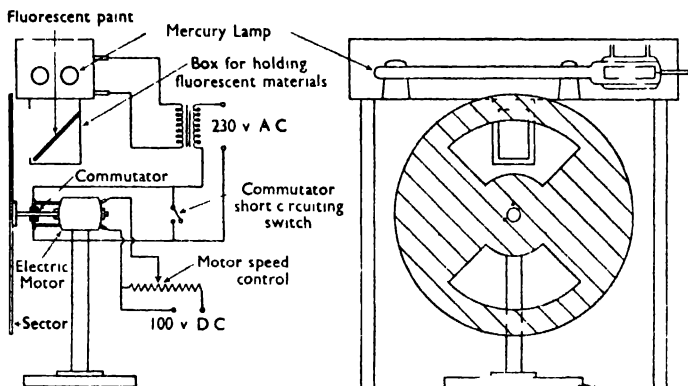


FIG 5   Sector Experiment

measured. Before it is possible to go very far with the analysis of a reaction it is necessary to know the concentration of the active species controlling the rate. This concentration, the average lifetime and the rate of production are related in the following way.

*Concentration = Rate of production × Average lifetime.*

Hence if the rate of production is determined, say, in a photochemical reaction from the intensity of the absorbed light, then the concentration can be computed provided the lifetime is known. Some means are therefore required whereby the decrease in concentration of the active species with time

may be measured, when the rate of production is suddenly decreased to zero. On account of the short lifetime it is impossible to measure the chemical effect during one such decay period. Some method of adding together the effect of a large number of decay periods must therefore be devised. This may be done with a rotating sector. The principle of the method can be most conveniently demonstrated by the use of fluorescent materials, and the problem then is: How may the lifetime for the decay of fluorescence be measured? The sector is set up as shown in fig. 5 and driven by a universal motor controlled by a potentiometer. The fluorescent substance, supported on black cardboard, is placed immediately behind the sector in a box attached to a mercury-neon cold cathode lamp. The brushes of the commutator on the motor shaft are so positioned that the mercury lamp is switched on while the fluorescent is screened by the sector. Just before the sector exposes the fluorescent solid the lamp is switched off. If, therefore, the decay period of the fluorescence is smaller than the interval between successive periods of illumination, then the intensity of fluorescence will be small. On the other hand, when the decay period is longer the fluorescence will be bright. In other words, the intensity will increase with speed of the sector. The effect is most easily shown by running the sector at a constant speed of a few hundred revolutions per minute and inserting into the box a series of substances of different decay periods, such as calcium tungstate, $< 10^{-6}$ sec., willemite, $10^{-1}$ sec. and less depending on the method of preparation, and calcium sulphide which has a decay period of several seconds. It is convenient to short circuit the commutator so that the fluorescence may be shown before

commencing the experiment. With willemite the variation of intensity with rate can often be demonstrated if a suitable sample is obtained.

The mode of operation of the mercury lamp used for the above experiments may be strikingly shown by the discharge tube shown in fig. 6. It is made of glass and fitted with iron electrodes such as are used for neon signs. On the left side some powdered willemite is fused on to the inner surface of the tube. One small drop of mercury is run into the side tube and neon admitted to a pressure
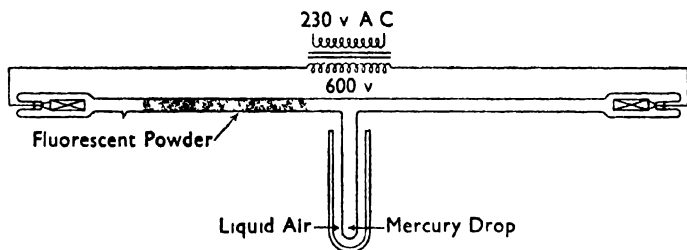


Fig 6. Mercury-Neon Discharge Tube

of a few millimetres. With liquid air on the side tube, the discharge has the characteristic colour associated with neon* but when the liquid air is withdrawn the mercury vapour enters the discharge tube and the colour gradually changes from red to blue. As the mercury reaches the fluorescent powder the latter radiates a strong green fluorescence showing that it is only when mercury is present that sufficient ultraviolet light is produced to excite the fluorescence of the willemite. The mercury lamp itself is, of course, fabricated of fused silica to transmit this radiation. Incidentally, the experiment demonstrates the rate of diffusion of

* The neon should be passed through a liquid air trap to free it from mercury vapour, as the removal of the vapour by placing the side tube in liquid air is sometimes a very slow process, presumably owing to its absorption by the iron electrodes.

mercury through neon. On re-applying liquid air,
the discharge once more becomes red.

The photochemical method of producing acti-
vated molecules and so inducing reaction to occur
has the great merit that the conditions in the system
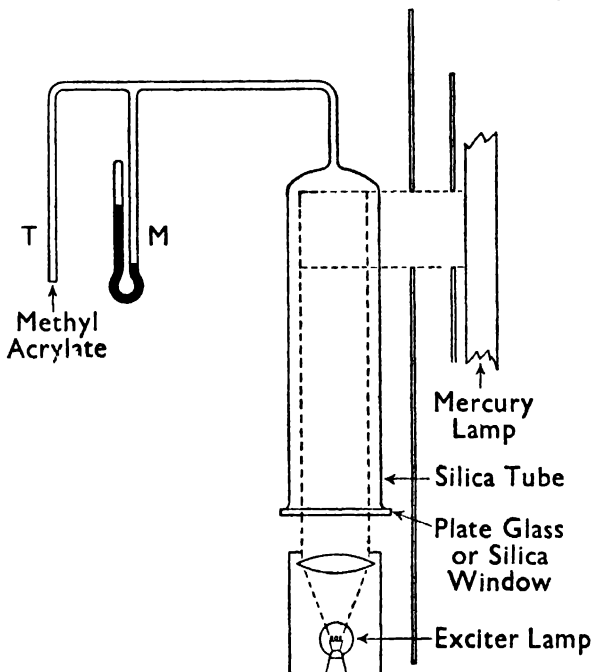


FIG 7. Apparatus for polymerization of methyl acrylate.

may be changed independently of the source of
excitation. This measure of control is much used
in unravelling complex reactions when a whole
series of consecutive processes have to be separated
into their components. This is particularly necessary
in the study of the polymerization of simple
molecules to synthetic resins, rubbers and similar
products. Space will not permit a discussion of these
matters, but the apparatus shown in fig. 7 may be

used to demonstrate how rapid polymerization may be in the gas phase. In fact, the molecules grow so big that they may be made visible. From the side tube T, methyl acrylate vapour* is allowed into the silica reaction vessel to a pressure of about 60 mm. as registered by the mercury manometer M. An intense beam of white light is projected along the axis of the tube from a small projection lamp. When the mercury lamp (similar to that used in the previous experiments) is switched on a cloud of polymer forms in a few seconds. Ultimately the molecules aggregate to larger particles and settle out of the vapour. If liquid air is applied to the side tube the vapour soon sweeps the polymer out of the reaction vessel and the Tyndall cone disappears.

[H.W.M.]

## APPENDIX ON THE DETAILS OF THE CONSTRUCTION OF THE POTENTIAL ENERGY MODEL

Some notes may be added about the construction of the model described on p. 370. The runway consists of a length of brass strip $\frac{1}{2}$ in. wide. It is confined within two sheets of window glass by a wooden frame and is self-supporting. At the top of the frame there is a hole to admit $\frac{3}{8}$ in. steel balls and at the left side there is a small cup to catch the steel balls. A piece of white cardboard is fixed behind the model with the numbers and lettering about 4 in. high. At the right side there is a brass plate drilled with three holes and fitted with four screwed studs upon which the gun rests. The two upper holes lead to two $\frac{1}{2}$ in. phosphor bronze strips either of which may be allowed to rest on the

* The methyl acrylate should be thoroughly freed from air before finally evacuating the apparatus. In this state it may be preserved indefinitely in a $CO_2$ bath at 80°C.

runway by means of two wires anchored at the top of the frame. The detailed construction of the gun is shown in fig. 8. The barrel carries plunger P attached to a knob K and controlled by a spring S. B is screwed into a plate C which is held to the apparatus by two screws T. A short length of tube D carries a rod upon which is hung a brass
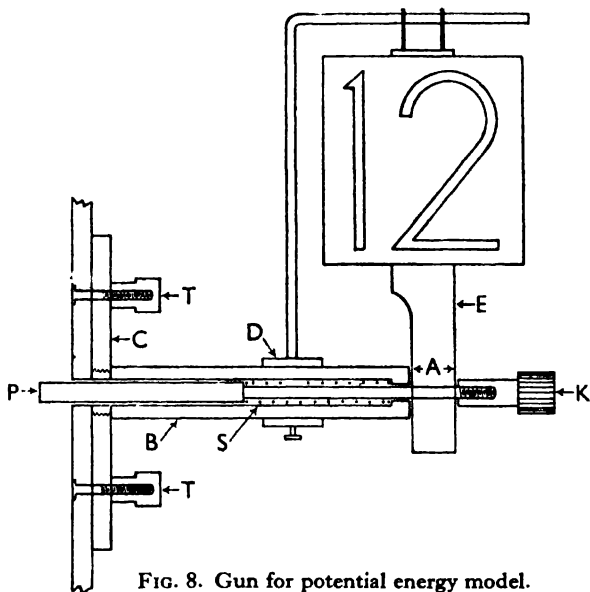


FIG. 8. Gun for potential energy model.

strip E carrying a square of cardboard bearing the energy imparted to the ball by the gun. The end of E is milled down to a distance A. This part of the strip is inserted between the plunger P and the knob K. The gun is fired by sharply flicking E from its position between B and K. The distance and/or the strength of the spring are so adjusted that the ball just passes over the brass strip according to the position of the gun on the studs of the apparatus.

[H.W.M.]

## Friday, November 11, 1938

SIR ROBERT ROBERTSON, K.B.E., M.A., F.R.S.
*Treasurer and Vice-President, in the Chair*

SIR WILLIAM BRAGG, O.M., K.B.E, M.A., D.Sc.,
Pres.R.S., M.R.I.
*Fullerian Professor of Chemistry*

## COMBINATION TONES IN SOUND AND LIGHT

WHEN two notes are sounded together the listener may hear in addition other notes which are not overtones of either of the originals. They are combination tones, whose frequencies are linear algebraic functions of the primaries. Let us denote the latter by $m$ and $n$. The most powerful combination tone has a frequency equal to the difference between $m$ and $n$ : a note of frequency $m + n$ is much more difficult to observe. There may be notes of other frequencies $m + 2n, 2m + n$ . . . . but these are weaker still. A full description of the effect is given by Jeans in " Science and Music ", p. 287 *et seq*. It is possible to make experimental arrangements by which the existence of combination tones may be plainly demonstrated. Some of these arrangements are described in the present note.

The subject has had a curious history. Observation must be somewhat close and careful if combination tones are to be heard ; this is no doubt the reason why no one described them before Sorge in 1745 and Tartini in 1754. Helmholtz was greatly interested in them and discussed them at length in his famous work on "Sensations of Tone". Koenig strongly opposed

the theory which Helmholtz put forward in explanation. The point at issue was whether or no the extra tones were subjective or objective. Helmholtz claimed to have shown by experiment that they caused a response in tuned resonators, Koenig asserted that they had their origin only in the listener's ear. A number of other experimenters joined in the discussion. Lord Rayleigh, Waetzman, Edser and Rucker supported Helmholtz, while Preyer and Bosanquet sided with Koenig. Ellis in the second edition of his translation of Helmholtz' book, devoted a long chapter to a reasoned comparison of the two theories.

In view of the vigour of the disputation it is strange to observe that, when certain conditions are satisfied, combination tones must have a real objective existence in the sense that resonators can be tuned to them. It is also true that, as Helmholtz pointed out, they can be generated within the ear, on account of the ear's peculiar structure. Certain preliminary conditions must be satisfied if the tones are to be objective, and it would seem that the long controversy was maintained by failure to realize those conditions in all cases.

The discussion at last died away. It was not even referred to when fifty years later an exactly analogous problem was met with in wireless telephony and the discussion was repeated in full. The so-called side bands are combinations of the carrier wave and the audio wave, and in the sense already described, have just as real an existence.* Still later, the phenomenon has been found again in the peculiar scattering of light known as the Raman effect.

*Hazel. Phil. Mag. xix. 103. (1935).

The underlying principle is very simple. When vibrations of two different frequencies are imposed upon a medium, as for example when two tuning forks are sounded, the chord consists only of the imposed frequencies and there are no combination tones in the air. But if during the production of sound the amplitude of one source depends at every instant upon the amplitude of the other the
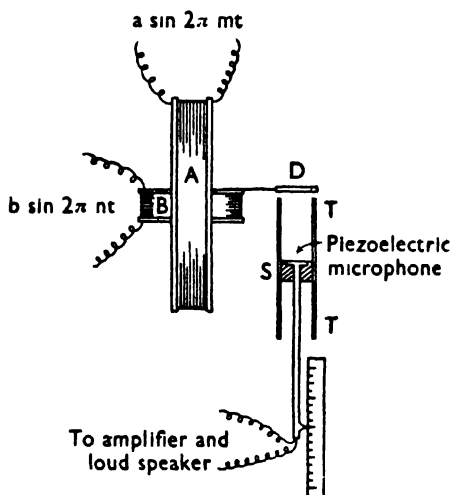


FIG. 1

combinational tones appear. In the first case the disturbance in the medium may be represented by

$$a \sin 2\pi mt + b \sin 2\pi nt$$

In the second there is sure to be a term of the form:

$$2c . \sin 2\pi mt \times \sin 2\pi nt$$

which is equal to

$$c . \cos 2\pi (m-n)t - c . \cos 2\pi(m + n)t$$

and the combinational tones have an objective existence.

This conclusion is readily illustrated by experiment. In the diagram (Fig. 1) A represents a coil seen edgeways, through which either a steady current $a$, or an alternating current $a \sin 2\pi mt$ can be made to run. B is a smaller coil at right angles to A ; the centres of the two coils coincide. Through the second coil, either a steady current $b$ or an alternating current $b \sin 2\pi nt$ may be made to run. The second coil can turn about a vertical axis which is perpendicular to the plane of the diagram. T is a cylindrical tube in which a slider S is mounted ; the slider contains a piezo-electric microphone connected to a loud speaker. When B vibrates about its axis, which it must do if there are currents in the coils and one or both are alternating, the disc D flutters before the mouth of the tube. The tube, as a resonator, can be tuned by moving the slider.

In an actual experiment the current of frequency $m = 250$ was taken from a high frequency generator, that of frequency $n = 50$ was drawn from the lighting circuit. When the current in A was alternating and that in B was steady, the resonator responded to the 250 note. When the current in B was alternating, the response to the 250 note became much weaker but the resonator could be made to respond strongly to either 200 or 300. The torque on the B coil was proportional to

$$\sin 2\pi . 250 . t \times \sin 2\pi . 50 . t$$

that is to $\quad \cos 2\pi . 200 . t - \cos 2\pi . 300 . t$

The difference tone should be the stronger, other things being equal, because the equation of motion of the B coil may be written

$$K . \ddot{\theta} = -c \left\{ \cos 2\pi(m-n)t - \cos 2\pi(m+n)t \right\}$$

where $\theta$ is the angular displacement and $K$ is the moment of inertia of the coil and $c$ is a constant; so that

$$K.\theta = \frac{c}{4\pi^2} \left\{ \frac{\cos 2\pi(m-n)t}{(m-n)^2} - \frac{\cos 2\pi(m+n)t}{(m+n)^2} \right\}$$

The movement of the disc at the mouth of the resonator can therefore be analysed into two harmonic terms, one of them having a frequency $m-n$ and an amplitude proportional to $1/(m-n)^2$,
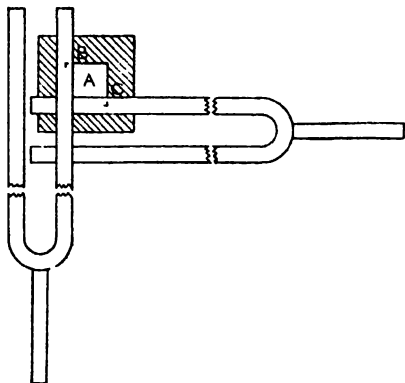


FIG 2

the other a frequency $m+n$ and an amplitude proportional to $1/(m+n)^2$. The response of the piezo-electric microphone, and the other electric and acoustic transformations of the energy which take place before the sound leaves the loud-speaker, will alter the relative intensities of the two sounds and will even introduce fresh combination tones ; but the main effect is so pronounced that the verification of the theory is quite clear.

An effect depending on the multiplication of harmonic terms can also be shown in a second way. (Fig. 2.) The light which passes through a small

square aperture in a screen is brought to a focus at a place marked A in the diagram. It is there partly intercepted by two tuning forks which are at right angles to one another. The beam of light is thereby restricted to a rectangle of which the sides may be denoted when the forks are vibrating by

$$a + b \sin 2\pi mt \text{ and } c + d \sin 2\pi nt.$$

The intensity of the light which passes on is therefore proportional to the product of these two expressions ; and this product contains the term $bd \sin 2\pi mt. \sin 2\pi nt$. The light falls on a photocell and produces an effect which is transformed into sound. The latter can be analysed by the piezo-electric resonator used in the first experiment. It is found to contain the frequencies $m$, $n$, $m-n$ and $m+n$.

If the two tuning forks are placed parallel to each other the intensity of the transmitted light is proportional to $a + b + c \sin 2\pi mt + d \sin 2\pi nt$ and there are no combination tones due to the optical arrangement. There may be, however, faint combination tones due to the subsequent treatment of the energy in the detecting apparatus.
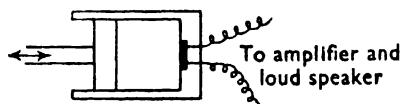


To amplifier and loud speaker

FIG. 3

The way in which a detecting apparatus can generate accessory tones may be illustrated by the following example. (Fig. 3.) Suppose that a piezo-electric cell is inserted at the end of a short cylinder, and that a piston executes a harmonic motion, so that the volume of the air may be put equal to $v = v_0 + v' \sin 2\pi mt$. Then the pressure

$p$, supposing the motion to be isothermal (though it makes no difference to the general result if it be taken as adiabatic) is given by $pv = c$, a constant.

Therefore
$$p = \frac{c}{v_0 + v''\sin 2\pi mt}$$

$$= \frac{c}{v_0}\left(1 + \frac{v'\sin 2\pi mt}{v_0}\right)^{-1}$$

and this when expanded shows overtones, although the piston itself moves in a simple harmonic motion.

If there are two pistons operating on the one chamber

$$p = \frac{c}{v_0}\left\{1 + \frac{v'\sin 2\pi mt}{v_0} + \frac{v''\sin 2\pi nt}{v_0}\right\}^{-1}$$

and this when expanded shows a system of combination tones.

The variations of pressure can be converted into a complex sound in which the tones appear.

When a tuning fork is held before the mouth of the resonating tube used in the first experiment, the octave overtone is readily found.

When two notes are sounded on the harmonium or organ, combination tones are generally heard, though some people find a difficulty, especially at first. The combination tones of a harmonium at the Royal Institution are greatly intensified when a microphone is placed on the case of the instrument and connected to a loud speaker. The combination tones of harmonium and organ would seem to be due to modulation, as in the simple instance just given. The pistons of the model are in the actuality replaced by the to and fro surges of air from the sounding pipes or reeds. It seems that the combination tones are of greater importance in the

vibrations of the woodwork of the chest than in the air outside.

In the radio transmission of sound modulation is a necessity. Sound waves can be converted directly into electrical oscillations of similar frequency, but it is practically impossible to put into such slow oscillations sufficient energy for
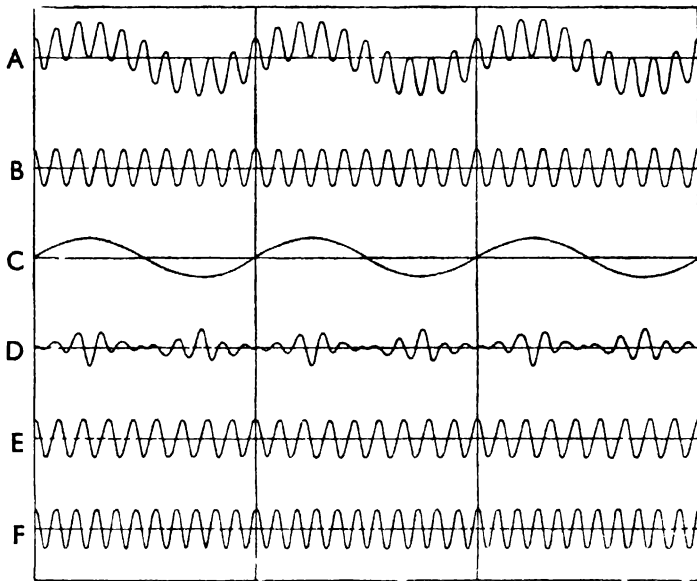


FIG. 4   The explanation is given in the text.

effective transmission. Hence a carrier wave of high frequency is employed which is modulated by sound waves and carries to the receiving station the impressions which have been made upon it. The form of the modulated wave can be represented in the customary way by such a figure as the curve D in Fig. 4, which may be described —approximately only—as a disturbance having a periodicity equal to that of the carrier wave, and an

amplitude which varies in accordance with the impressions made by sound at the transmitting end. Strictly speaking the disturbance is the sum of a number of different waves. One of them has the frequency of the carrier wave, and is as a matter of fact useless. Others are combination tones and these are the real carriers, and there are other derivatives present in relatively weak amount.

In the accompanying diagram (Fig. 4) the harmonic curve B may be taken to represent the carrier wave, that is to say the ether wave due to the oscillator of the transmitting station, and C the ether wave into which a harmonic sound is transmitted. The frequency of B is ten times that of C; in practice the factor is very much larger. If the disturbances were merely added as represented by the curve A, the two would be quite independent in their effects at a receiving station: the wave B would be detected just as if there were no wave C, and would convey nothing. The wave C would be too faint for perception. But if by proper manipulation at the sending station a wave is produced which is the continuous product of B and C, as shown in the curve D, then the effect of C on B is detected at the receiving station, that is to say the sound signal is received.

The curves E and F are harmonic curves of frequencies nine and eleven times that of C, being the difference and sum respectively of the frequencies of B and C; and D can also be obtained by adding E and F together. The receiving station can detect these two separately. They are combination tones. Technically they are known as 'side-bands'.

The Raman effect is based on the same principle. The incident light of definite frequency excites a forced vibration of the same frequency upon the

molecule which it meets. If the magnitude of the response to the impressed electric force depends upon some quantity in the molecule which is undergoing harmonic oscillations, as for example some relative oscillation of the atoms in the molecule, modulation occurs. Lines representing combination tones are found on each side of the line representing the original frequency in the spectrum of the scattered light.

[W.H.B.]

# Friday, December 9, 1938

SIR WILLIAM LARKE, K.B.E.
*Manager and Vice-President, in the Chair*

IRVING LANGMUIR, M.A., Ph.D., D.Sc.,
LL.D., Hon. M.R.I., For.Mem.R.S.
*Associate Director of the Research Laboratories of
the General Electric Company, Schenectady, N.Y.*

# THE PROPERTIES AND STRUCTURE
# OF PROTEIN FILMS

THE proteins form a class of chemical compounds that make up an essential part of all living organisms. They exist not only in the tissues and muscles as fundamental constituents, but they form also the blood serum and the various enzymes that play such an important part in the metabolism of animals. Pepsin and trypsin, for example, take part in the digestive processes; insulin helps to regulate the sugar metabolism. The toxins which are active agents in many diseases are proteins, as are also the anti-toxins that the body makes to combat disease.

Within recent years it has been found that the viruses, which were long thought to be living organisms, are proteins. Whether or not a protein itself can be endowed with life is perhaps a matter of definition, but certainly no known class of chemical substances is more intimately connected with the phenomena of life than the proteins.

A very remarkable characteristic of proteins, which is not encountered among other chemical substances, is that they show the phenomenon of denaturation when subjected to relatively mild heat or other influences too weak to produce marked changes in other substances. Most proteins have a

number of very specific properties; for example, diphtheria toxin is the only substance that produces symptoms of diphtheria and diphtheria antitoxin is the only protein that will neutralize the effect of this toxin. Besides the particular properties for which a given protein seems to have been designed, each protein has a specific antigenic power when introduced as a foreign protein into an animal, to cause the animal to generate an antibody (also a protein) which, under the proper conditions, can neutralize the effect of that particular protein.

All of these specific properties are lost when a protein is denatured. Heating the protein to 65° C. in the presence of water for 5 minutes is usually sufficient to denature the protein. If a bottle is half-filled with a dilute solution of a protein and is then vigorously shaken for 5 minutes the protein loses its specific activities. This is not an oxidation process for it occurs even if the bottle is filled with nitrogen.

Many proteins crystallize in cubic or hexagonal crystals, forms of high symmetry, and studies of these crystals by X-rays have shown that this high symmetry resides in the molecule itself.

There is thus much to indicate that the proteins have highly organized structures, and that the properties of these substances depend to a much greater degree upon details of structure than is usual among chemical compounds.

When proteins are boiled with acid they break down into compounds known as amino-acids which have the general formula

$$H_2N\text{-}CHR\text{-}COOH$$

where R represents a group of atoms known as the side chain. There are somewhat over twenty differ-

ent amino-acids that can be obtained from proteins. These differ only in having different groups of atoms in the side chain R. There is much chemical evidence that during the early stages of the breakdown of the proteins long polypeptide chains are formed which consist of amino-acid residues

-HN-CHR-CO-

arranged as links in a long chain. Each of these residues is obtained from the corresponding amino-acid by the splitting off of one water molecule.

Within recent years there has been much discussion as to the structure of the undenatured proteins (native proteins) and many have supposed that they consist of long polypeptide chains and that the particular sequence of the amino-acids determines the specific nature of the protein. Others, however, of whom I am one, believe that the evidence for the high organization in the protein molecule, which leads to its specific properties and highly symmetrical form, requires a much closer knitting together of the molecule than would be given by a merely linear arrangement in a chain. Dr. Wrinch has proposed a particular fabric-like structure which is folded into a definite polyhedron.

Much work has been done in determining the molecular weights of the proteins. A large number of proteins such as egg albumin, pepsin and insulin have molecular weights close to 36,000. Most of the other proteins have weights which are multiples of 18,000, some of them as high as 100,000,000. There is much to indicate that each protein is a very definite chemical substance having a constant molecular weight, although in some cases the molecules may aggregate into larger molecules.

The problem of the structure of protein molecules is one which is being worked on very ener-

getically at present, and we may expect within very few years that enormous progress will result which will have a profound bearing on biological problems and will lead us to a clearer understanding of the simplest forms of life.

Within recent years methods have been developed for studying the physical and chemical properties of single layers of molecules spread upon the surface of water. These methods provide very simple ways of measuring the sizes and shapes of molecules and throw a great deal of light on the nature of the forces that act between different parts of the surfaces of molecules—forces which often cause the molecules to be oriented on the surface. The majority of proteins can also be spread out in thin films on water having thicknesses of about $10^{-7}$ cm. and thus the methods of surface chemistry are applicable to the study of these substances. This evening I propose to describe some of the elementary phenomena which are observable with single layers of molecules (monolayers) on water surfaces and will then show how these methods may be applied to the study of proteins.

## CAUSE OF SPREADING OF FILMS ON WATER

Pure saturated hydrocarbons, such as liquid petrolatum or hexadecane, $C_{16}H_{34}$, when applied in the form of drops to the surface of clean water in a long trough, remain on the surface as lens-shaped drops which may be several mm. thick, but show no tendency to spread out as a thin film over the surface of the water. These hydrocarbons are also insoluble in water. The pure hydrocarbons thus have no particular affinity for water and the molecules may therefore be described as *hydrophobic*.

Organic substances such as sugar which contain

a large number of hydroxyl groups, -OH, are readily soluble in water. Ethyl alcohol, $C_2H_5OH$, is completely soluble in water, whereas the corresponding hydrocarbon ethane, $C_2H_6$, is a gas that is relatively insoluble in water. Clearly, therefore, the -OH group has a marked affinity for water and may therefore be called *hydrophilic*. Another important hydrophilic group is the carboxyl group, -COOH, which is present in nearly all organic acids. We can thus draw the general conclusion that substances whose molecules are covered with hydrophilic groups dissolve readily in water and those which contain only hydrophobic groups will be practically insoluble.

Let us now consider what happens when we deal with substances whose molecules are mostly hydrophobic but contain a single strongly hydrophilic group. Consider, for example, hexadecane in which one of the $-CH_3$ groups at the end of the long hydrocarbon chain is replaced by a carboxyl group, -COOH, forming the compound known as palmitic acid. It is hardly to be expected that a single group of this kind can force the whole of the long hydrocarbon chain to dissolve in water. What actually happens is that the molecules of palmitic acid spread out over the surface so that each hydrophilic carboxyl group, or head, comes into contact with the water without requiring that the hydrophobic parts, or tails, should separate from one another; thus both the hydrophilic and hydrophobic tendencies of the different parts of the surface of the molecule are satisfied.

If there is a limited surface of water and a surplus of palmitic acid, the heads crowd into a surface layer so that the hydrocarbon chains or tails, which have a cross section about the same as that

of the carboxyl group, are vertically oriented so that they form a tightly packed layer of hydrocarbon molecules. The thickness of the layer of palmitic acid thus measures the length of the hydrocarbon chain and the area per molecule gives the cross-section of the hydrocarbon chain. Another fatty acid having a longer hydrocarbon chain, such as cerotic acid (26 carbon atoms instead of 16) will give the same area per molecule but the thickness of the film will increase about in proportion to the number of carbon atoms. There is no tendency to form a second layer of molecules because the heads of the molecules of a second layer cannot come into contact with water.

This theory of the cause of the spreading of oils is based upon the postulate that the forces between molecules are of such short range that they act only between the surfaces of adjacent molecules which are in contact.

To produce a film of palmitic acid on water it is merely necessary to dissolve a little of the solid in a volatile solvent such as benzene (about 1 part in 300), and place a drop of this solution upon a clean surface of water in a trough filled to the brim. The method of cleaning the surface of the water, which is very important in experiments of this kind, is merely to place a barrier across the trough, supported on its edges, and to sweep this barrier the whole length of the trough so as to produce a fresh surface of water behind the barrier. The palmitic acid spreads out to form an invisible film of a thickness of about $22 \times 10^{-8}$ cm., or as we may more conveniently say 25A. where we use A. to represent the Ångstrom, $10^{-8}$ cm. We can detect the presence of the film by observing its effect in pushing strips of paper or dust particles ahead of it as it advances

over the surface, even to the far end of the tray. In fact, if enough of the palmitic acid is added in this way it exerts a pressure against a barrier at the far end which may be measured by a simple surface balance. For this purpose it is merely necessary to have a floating barrier connected at its ends to the edges of the tray by waxed silk threads which prevent the escape of the monolayer on to the surface of the water behind the barrier. To hold the barrier in a definite position against the pressure exerted by the monolayer requires the application of a force which can be measured. A monolayer on the surface of the trough can be compressed by moving a barrier behind it and the surface pressure that is thus generated can be measured by the surface balance.

Just as we can study a 3-dimensional gas by putting it in a cylinder and applying pressure to it by a piston, observing the change of volume produced, so we can compress a monolayer in a trough by a moving barrier (2-dimensional piston) until the force indicated by the balance has reached any desired value, and can then observe the area of the monolayer. We find in this way that a palmitic acid monolayer behaves as a 2-dimensional liquid, as it is relatively incompressible and does not tend to increase indefinitely in area when the pressure is released.

The reason that the palmitic acid molecules act as a 2-dimensional liquid is that the hydrocarbon chains tend to stick together just as if the carboxyl groups were not present (in which case the hydrocarbon would draw itself into a lens on the surface).

If we take a fatty acid having a much shorter chain than palmitic acid we find that the adhesion

between the chains becomes so small that the molecules are able to separate from one another on the surface and thus act as a true 2-dimensional gas. However, the shortening of the chain also brings about an increased solubility in the underlying water so that we can no longer use a surface balance and moving barriers to detect the monolayer on the surface. There are other methods, however, involving the measurement of surface tension, which enable us to study these soluble monolayers and to prove that they behave as 2-dimensional gases.

We say that the palmitic acid monolayer behaves as a liquid rather than a solid because if we place dust particles on the surface and blow upon them we see that these move freely over the surface. If we introduce into the water, however, certain salts such as aluminium chloride, the dust particles act as though frozen into the surface. We can then say that the monolayer is solid, the molecules presumably being bound together by aluminium atoms which have combined with them. Monolayers can thus exist in the form of 2-dimensional gases, solids, or liquids.

There are many insoluble monolayers on water which form so-called expanded films. In these the hydrocarbon chains are not oriented parallel to one another but are irregularly arranged as in a liquid, one end of each molecule, however, having its hydrophobic group bound to the water surface, although free to move within this surface. The hydrocarbon parts of the molecule thus constitute a 3-dimensional liquid or interstratum which is bounded on its upper surface by an interface like that between any hydrocarbon liquid and air. The lower interface, between the hydrocarbon and the

water, contains the hydrophilic heads which move over the surface as a 2-dimensional gas. This gas cannot expand indefinitely, however, because as it increases in area the expansive force decreases until finally it is only able to balance the tendency of the hydrocarbon liquid to draw into a lens. The expanded films of fatty acid thus spread out to 50 or 60 sq. A. whereas the condensed films of vertically oriented molecules give an area per molecule of 20 sq. A.

## THE USE OF INDICATOR OIL TO RENDER MONOLAYERS VISIBLE

A pure petroleum lubricating oil does not ordinarily spread to a thin film on the surface of water because it does not contain hydrophilic groups. If it is heated on a hot plate until it smokes it gradually oxidizes and hydrophilic groups are produced. The oil thus spreads to thinner and thinner films as the oxidation proceeds. By stopping the oxidation at the proper stage, or by mixing oils of two different degrees of oxidation, one can obtain an oil which spreads to give the first order interference colours on Newton's scale. The colours are produced by the interference of light as it is reflected from the top and from the bottom surfaces of the oil film. The thinnest film that gives good colours is one which corresponds to about one-quarter wave length of visible light (about 1000 A.)

A drop of such indicator oil on the surface of water gives a film of a uniform colour. If we compress this between barriers we make it thicker and thereby change the colour. The colour can thus be made to serve as a measure of the degree of compression or surface pressure.

An amount of indicator oil sufficient to cover

only about half the surface of the water in the tray does not change the surface tension of the water. If we now apply to the centre of this oil patch a substance which spreads to form an insoluble monolayer this forms a black spot or disc in the centre of the oil patch. To observe these effects most easily the bottom of the tray should be painted black.

If the monolayer is liquid the application of the droplet of indicator oil gives a circular coloured disc. With a solid film, however, the indicator oil tears the monolayer into jagged cracks, forming a star-like figure or expansion pattern.

## PROTEIN MONOLAYERS

If a small fragment of insulin or egg albumin or other water soluble protein is brought into contact with water in the centre of a patch of indicator oil, the protein spreads out to form a monolayer which gives a pattern of characteristic shape. Insulin and pepsin give circular patterns, while egg albumin gives a pattern of very irregular outline resembling the profile of cumulus clouds on a summer evening. If a droplet of indicator oil is now applied in the centre of the protein monolayer this spreads out into a characteristic pattern. For example, with insulin the indicator oil forms a circle, whereas with pepsin and with egg albumin it forms star-shaped figures.

The amounts of protein necessary to produce these monolayers is surprisingly small, 1 mg. being enough to cover a sq. meter or only $10^{-7}$ g. will give 1 sq.cm., enough to produce a distinct pattern. These very characteristic patterns serve as an easy method of identifying certain native proteins. If the proteins are then subjected to some treatment that causes a gradual denaturation, one can follow

the changes by the expansion patterns that are produced. One frequently finds that different types of denaturation produce different types of expansion patterns.

Although the proteins themselves are frequently very soluble in water the monolayers formed from them are extraordinarily insoluble. This shows that a radical change of structure has occurred during the spreading of the protein. Substances whose molecules are wholly hydrophilic have no tendency to spread on the surface of water. The very high solubility of certain proteins thus proves the absence of hydrophobic groups on the outside surface of their molecules, yet the spreading to form insoluble monolayers proves that the molecule when unfolded possesses large numbers of hydrophobic groups. The native protein must, therefore, have a cage-like structure (such as that postulated in Dr. Wrinch's cyclol theory), in which the hydrophobic groups are normally buried in the interior. When these molecules come into contact with the surface of water the hydrophobic groups are drawn to the water surface causing a tearing open or unfolding of the cage. The insolubility of the monolayers thus results from the fact that the tearing open of the highly organized cage is an irreversible process.

Protein monolayers are extremely easily and reversibly compressible. They may, in fact, be compressed to one-fifth area and yet be able to return immediately to the original area if the compression is removed. The properties of the protein monolayers seem to prove that the monolayer is a kind of duplex film which consists of polypeptide chains

of amino-acids, attached to the air-water interface at intervals along their lengths by hydrophobic groups. The hydrophilic parts of the polypeptide chains remain in contact with and are surrounded by water, but are not free to go far from the surface because they form parts of chains which are still attached to the surface. When the film is compressed there may not be room enough for all the hydrophobic groups on the surface so that some are driven down into the water just below the surface. Since these hydrophobic groups, like oil drops under water, tend to coalesce or adhere to each other, they are at least partly responsible for the viscosity and rigidity of the monolayers shown by the expansion patterns.

If a small metal disc, 1 inch in diameter, is suspended in a horizontal position by a fine wire and is provided with a dumbbell-shaped weight to increase its moment of inertia, it will oscillate after being set into rotation about its axis. If such an oscillating disc be lowered into contact with a monolayer of protein on a water surface, the damping of the oscillations provides a means of measuring quantitatively the 2-dimensional viscosity of the monolayers. The measurements are very reproducible and highly characteristic of the different proteins, some proteins giving viscosities more than 10,000 times as great as others. I believe that these viscosities will help to throw light upon the nature of the cross-linkage between polypeptide chains; they should also serve as an accurate method of following changes produced by denaturation.

### DEPOSITED PROTEIN MONOLAYERS

Protein monolayers can be transferred by a simple deposition process from a water surface to the sur-

face of a metallic plate which is dipped into the water. If this plate has previously been coated with the proper number of layers of barium stearate, built up by Dr. Blodgett's method, brilliant interference colours are produced which serve for an accurate determination of the thickness of the deposited monolayer. By this method it is possible actually to see the change of colour produced by a single monolayer of the protein. By using monochromatic light, such as that given by a sodium vapour lamp, the thickness of a single monolayer can be measured to within about 2 A. It is also possible to deposit in succession many hundreds of monolayers, forming a multilayer. In this way the thickness per layer can be determined by optical measurements with great accuracy. Experiments prove that in the formation of these deposited layers whole monolayers can overturn, involving an interchange of the positions of the hydrophilic and hydrophobic groups.

Metallic plates covered with barium stearate can be so treated with a thorium nitrate solution that the film becomes hydrophilic or wettable by water, and it is then capable of adsorbing proteins directly from water solution. The plate can then be washed and dried and the change of colour measures the thickness of the layer of protein adsorbed from the solution. It is then possible to bring the plate into contact with another solution containing a protein which reacts with the first one deposited. This reaction may cause the second one to be deposited on the first or may in some cases cause the first to be removed by the second. This technique thus makes possible the study of the reactions between proteins in a new way. The results have already indicated that the reactions are extremely specific

for different proteins and are very sensitive to changes of conditions which are known to produce important biological effects.

The techniques for the study of proteins that are opened up by these methods of surface chemistry will, I hope, prove to be of value to the biologist in the work he is doing on the nature of the specific activities of proteins.

[I.L.]

Friday, January 27th, 1939

SIR ROBERT ROBERTSON, K.B.E., M.A., F.R.S.
*Treasurer and Vice-President, in the Chair*

J. D. BERNAL, M.A., F.R.S.
*Professor of Physics, Birkbeck College, University of London*

# STRUCTURE OF PROTEINS

THE structure of proteins is the major unsolved problem on the boundary of chemistry and biology to-day. We have not yet found the key to the problem, but in recent years a mass of new evidence and new lines of attack have enabled us to see it in a far more concrete and precise form and to have some hope that we are near to solving it.

It is hardly necessary to stress the importance of proteins, both biological and economic. Proteins are among the most, if not the most, characteristic chemical compounds which always accompany life and without which life is impossible. Much of our food and our clothes are protein products, and in cooking and preparing them mankind has used unconsciously the procedures of protein chemistry.

Proteins first appeared in science at the beginning of last century as rather vague and messy natural products containing carbon, hydrogen, oxygen and characteristically nitrogen. At the end of the century Fisher showed that they were composed largely if not entirely of amino-acids and suggested, though it was impossible to prove, that these were attached together through loss of water in a peptide link. This represents virtually all we know on the chemical side of the structure, as apart from the composition, of proteins, the advances of the present century have been nearly all of a physical nature.

In living organisms proteins can be divided into two groups according to their distribution. There are those which exist in solution both in the contents of the cell and in the fluids which bathe them. Among these are to be numbered a rapidly increasing collection of chemically active enzymes as well as other proteins, such as many albumins and globulins whose physiological function is not clear. Proteins of a different kind are to be found in the structural parts of the organisms, in the cell walls, in muscle fibres, in all forms of connective tissue, and external structures such as hair and horn. Intermediate between these two classes there seem to be the nucleo-proteins which form the connected structures of chromosomes, are found also in viruses and seem to be connected with processes of growth and reproduction. These biological differences mark out fundamental differences in physical character. The first group of proteins can, with appropriate technique, be made to yield crystals. They consist, therefore, of particles of definite and identical shape, and all other physical properties, solubility, osmosis and electro-phoresis confirm this picture. They have been called the globular, and more recently and perhaps more correctly the corpuscular proteins.[1] The proteins of the organic structures on the other hand are essentially fibrous in character and resemble other artificial and natural fibrous substances such as cellulose, rubber and plastics, they cannot be crystallized, contain no definite molecules, and may be called the fibrous proteins.

There exists, however, a very close connection between corpuscular and fibrous proteins. All corpuscular proteins are subject to a peculiar change when treated by a variety of agents, heat, strong

acids or alkalies, etc. Denaturation, as this change is called, leads, at any rate if it is violent enough, to an irreversible loss of solubility and crystallizing power; we are all familiar with it in the boiling of an egg. Now Astbury has shown that the denatured product is a typical fibrous protein.([2]) The plain conclusion from this is that the fibrous proteins are corpuscular proteins which have been still further aggregated and which have lost or have never attained their full degree of regularity. They have lost what we may call their long range order but preserved, as we shall see, much of their short range order.

The problem of protein structure can thus be seen to be divided in two; the first is that of the form and properties of the protein molecule, and the second that of its internal structure. Owing to the extreme instability of the protein molecule only the gentlest physical methods can be used, nevertheless three lines of attack, centrifugal, electrical and X-ray, have already led to great success. The most fundamental has been the ultra-centrifuge, particularly the work of Svedberg and his school.([3]) Protein molecules will sediment in sufficiently high centrifugal fields, and from their sedimentation constant it is possible to arrive at a fairly accurate value for the weight of the molecule, though there is some doubt as to whether what is measured is that of the complete protein molecule, because of its inevitable association with the solution in which it is suspended. The most striking discovery of Svedberg was that the weights thus obtained seemed to fall into definite classes which were multiples of each other. This suggests very strongly that all proteins are built from some common unit. What that unit is is more difficult to determine.

Originally taken as 35,000, proteins are now found of 17,000 and even of 10,000 molecular weight. Critical examination of the data moreover shows that there is no exact correspondence to certain weights, but rather a scattering of weights concentrated in certain regions. There is no doubt that in closely related proteins there are simple relationships. Both hæmoglobins and hæmocyanins can, for instance, be split reversibly into 2, 4, or 8 parts. But whether these relationships hold for all proteins is more doubtful; particularly as studies of viscosity have shown that proteins belonging to the same weight class, such as insulin and gliadin, may differ enormously in shape, the first being a flattened sphere and the other an elongated rod, and it is difficult to see what common physical structure they can have.

It has been claimed that this variation of weight of proteins of the same Svedberg class is to be expected on account of the known difference in amino-acid content. On the basis of existing analytic figures it is as yet impossible to say whether this is or is not sufficient to account for the observed discrepancy. The figure of 288 amino-acid residues in a 35,000 molecular weight class protein cannot be taken at present as more than an inspired guess.

The second approach to protein structure comes from the electric properties of their molecules. The work of Cohn, Tiselius and others gives us a picture of a protein molecule in solution as a particle covered with positive and negative charges due to the acid and basic nature of the amino-acid side-chains. The number of these groups depends on the condition of the medium. The protein molecule is, therefore, to a very high degree in its external relationship an ionic structure and probably carries
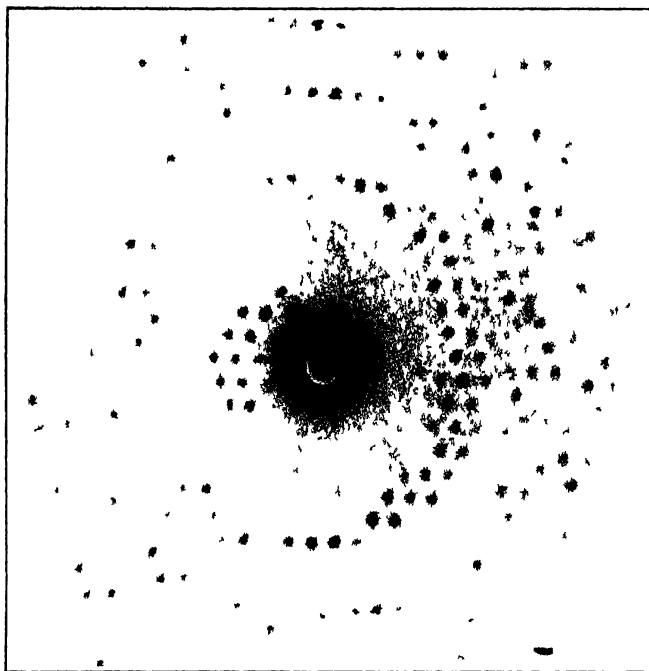
PLATE I    X-ray oscillation photograph of a crystal of horse
methæmoglobin showing (hol) zone, Perutz

with it an ionic atmosphere stretching out into the water in which it is dissolved.

The third line of attack is that of X-ray study. The beautiful crystalline forms exhibited by proteins, which have been known for at least a hundred years, were always a powerful attraction to X-ray crystallographers but until the last five years the instability of protein crystals and their small size had defeated their efforts. Since 1934,([1]) however, it has been possible to examine a number of typical protein crystals. This has been done by mounting them in small tubes in their mother liquor, for most, if not all, proteins suffer considerable breakdown if examined dry. From the beginning the results of this examination revealed important new facts. In the first place, the pictures yielded by protein crystals were of exceptional perfection. They showed large unit cells with a great wealth of reflections, (Plate 1) and these reflections were found even at comparatively high angles corresponding to such low spacings as 2A. This indicated that not only were the molecules of the proteins substantially identical in shape and size, but that they had identical and regular internal structures reaching right down to atomic dimensions.

From the size of the unit cells and their densities it was possible to compute the weight of matter in each repeat unit. This, however, does not immediately give the molecular weight for two reasons. In the first place the number of molecules in the cell is not known, and in the second it is difficult to determine how much of the cell weight is due to the protein and how much to the water in the cell. By measuring the loss of water on drying, it is however, possible to arrive at a figure for the—

possibly fictitious—dry weight, which can be compared with those obtained by centrifuge or chemical means. Here it is found that there is excellent agreement on the assumption that there are only a few, 2, 4, 8 or such molecules per cell.([5]) Thus the X-ray method furnishes an extremely accurate measure for the basic molecular weight of proteins though it cannot determine its multiplicity. It can give a higher but not a lower limit to the repeat units of the protein structure. Bermann([6]) and his co-workers claim to have found this lower limit by chemical methods. It is given by the smallest unit that contains an integral number of all the amino-acids estimated by analysis. But as Neuberger([7]) has pointed out, this method is not really reliable for high molecular weights owing to necessary inaccuracies in analytical technique. It suffers also from the intrinsic ambiguity as to what is to count as a protein molecule. Thus, for example, hæmoglobin was found by purely chemical means to have a molecular weight of 68,000 which agrees with that found by centrifuge. On the other hand, it is possible to split the hæmoglobin molecule into two, and the X-ray evidence necessitates either the existence of a twofold axis of symmetry in the molecule or that of two molecules of half the molecular weight. It may well be that many proteins are built of sub-units which, though of approximately equal weight, are not chemically identical and are more properly to be called molecular compounds than molecules.

The most striking feature of the cell measurements of crystalline proteins are the considerable changes which occur on drying, where in many cases shrinkage of nearly 50% occurs. But what is more remarkable is that the shrinkage is often

confined to one or two directions in the cell. There are two possible explanations of this fact. First, that the molecules of the protein are linked together in extremely loose aggregates which collapse on the removal of water, leaving the skeleton of molecules in a now more closely packed array. The other explanation is that the molecules are held apart by their ionic atmospheres due to the charges on them and contain, therefore, sheets of free water. It may be that both explanations hold, one more markedly for some crystals and the other for others. Thus for hæmoglobin, where there is a marked shrinkage from 55A. to 38A. in one direction, it is difficult to imagine an ionic atmosphere as the crystals are practically salt free and at an isoelectric point of pH 6.8. On the other hand, it is difficult to explain the remarkable properties of tobacco mosaic virus on any other hypothesis. This virus has long thin particles which have a tendency to set equidistant and parallel, even down to concentrations of 13% and probably down to 1.5%. It is difficult to imagine what other force other than that of ionic atmospheres can preserve this regularity. Quite recently we have shown that the equilibrium distance between the particles depends on pH and salt concentrations, varying for instance between 320A. at pH 7, and 206A. at pH 3.4. It should be possible to find a quantitative theory to explain these changes and, indeed, the beginning of such a theory has been made by Langmuir[8] and Levin.[9].

The general picture of the external character of protein molecules is beginning to be definite. The molecules are spheroidal bodies of dimensions varying from 30-100A., are covered with hydrophil groups bearing charges of both signs and probably

carry with them in solution an atmosphere of ions.

The problem of the internal structure of protein molecules is one of enormously greater difficulty. Any effective picture of protein structure must provide at the same time for the common character of all proteins as exemplified by their many chemical and physical similarities and for the highly specific nature of each protein type. It is reasonable to believe, though impossible to prove, that the first of these depends on some common arrangement of the amino-acids. The first hint of this comes from a study of the fibrous proteins. Here, owing to their technical importance and their much greater ease of handling, a very considerable advance has been made, due largely to the well-planned and persistent researches of Astbury. [10] As far as we know there are only two main types of fibrous proteins—the keratin-myosin type and the collagen-elastin type. Both of these exist in two forms, a folded elastic form ($\alpha$) and a fully extended form ($\beta$). Both show indubitable fibre structure, pointing to long polymerized molecules. While the precise arrangement of molecules in the collagen type is still unknown, the $\beta$ keratin type, which is also that of silk fibroin, seems to fit very well what would be expected from a straight peptide chain. Its characteristic is a strong repeat at every 3.5A., corresponding to the length of one amino-acid residue, and two other repeats at 10A. and 4.5A. respectively, corresponding to periodicities at right angles to the chain length. One of the most significant of Astbury's discoveries was that these two periodicities are also at right angles to each other. In other words, that the repeat unit of the fully extended protein has distinct characters in three dimensions. The significance of these repeats seems,

also beyond doubt, the 10A. spacing corresponding to the length of the side-chains of the amino-acid residues, and the 4.5A. to the so-called backbone spacing between the main chains themselves which are held together by the CO and NH groups. In the shorter $\alpha$ form the 10A. spacing remains, while the 4.5A. disappears, and this indicates that the chains are folded but folded definitely in a plane and not as in rubber in two dimensions. What the precise mechanism of this folding is remains obscure, but it is clearly of fundamental importance because it is probably very closely associated with the kind of folding that occurs in the corpuscular proteins.

The significance of these findings for the general protein problem is, firstly, that all soluble proteins hitherto examined give on denaturation a fibrous material which can be oriented and seems identical in basic structure with $\beta$ keratin, and secondly, that the actual change which a crystalline protein undergoes on denaturation must be very slight. This is shown by the close resemblance of powder photographs of crystalline proteins before and after denaturation, and also by the observations of Astbury, Dickinson and Bailey ([11]) that a single crystal of excelsin gives on partial denaturation fibres orientated in the direction of the crystal axes. Further, recent work by Perutz ([12]) has shown by an entirely different method, namely the study of the absorption spectra in different directions in crystals, that the position of the prostetic group in hæmoglobin is not altered by denaturation.

The evidence that the X-ray study of the crystalline proteins themselves provides for the elucidation of their structure is abundant, but it is extremely difficult to interpret. Photographs of

crystalline proteins show hundreds of spots and marked differences of intensities stretching right out to reflections corresponding to inter-atomic distances. Unfortunately, however, direct analysis of these photographs is rendered impossible by the fact that we can never know the phases of the reflections corresponding to the different spots. The ambiguity introduced in this way can only be removed by some physical artifice, such as the introduction of a heavy atom, or the observation of intensity changes on dehydration which have not hitherto been carried out in practice. For all but one of the proteins examined the task is made still more difficult by the presence of more than one Svedberg unit in the cell. The resulting X-ray pattern in such cases depends both on the position of the molecules in the cell and on their internal structure. Fortunately the one exception, insulin, has only one molecule per cell. An X-ray study of dry insulin crystals has been made by D. Crowfoot([13]) and the intensities of the reflecting planes determined. It is to be regretted that it was not possible to examine the crystals wet because in the dry crystal no planes of spacing less than 8A. are observed, and consequently no evidence exists as to the fine structure of the molecule. The evidence on the coarse structure is, however, striking enough, it can most easily be seen from a study of the Patterson projections and Patterson-Harker sections which Miss Crowfoot has computed for insulin. It should be explained that these projections do not represent the distribution of the density of scattering matter in the cell but merely the product of the densities at points separated by a constant vector. It is not surprising, therefore, that the interpretation of these diagrams has already

given rise to acute controversy. It seems natural to assume, though it is in fact an arbitrary simplification, that the peaks in the Patterson projection correspond to distances between a small number of high concentrations of scattering matter. Attempts have been made to reduce the analysis of the pattern to that of finding a pattern of points in space that will give maxima at the places observed. All these attempts have in fact failed. It is easy to find a large number of point patterns which will give the correct basal plane projection but none of these will at the same time fit the sections. The claims to have done so([14]) are based on the arbitrary selection of a certain number of vectors, and the agreement disappears if all the vectors are taken.([15]) The failure to find a point solution suggests that we are dealing with groups of size of the same order as their distance apart, and that the appearance of the enhancement of reflections at spacings 10A. and 4.5A. corresponds to an arrangement of side-chains or back bone spacings similar to those occurring in the fibrous proteins. On the basis of present X-ray knowledge it is clearly impossible to arrive at any detailed picture of protein structure. Indeed, up to now the chief value of the X-ray method has been to disprove such hypothetical structures as have been put forward. It is however possible to formulate in a broad way some of the possible modes of arrangement and to suggest working hypotheses as a guide to future work.

The formal problems of structure of the protein molecules are:—

*a.* What is the nature of the link between amino-acid residues?

*b.* Does the linkage run through the whole of the

molecule or only through parts of it? In other words is the protein one unit held together by primary valence forces or is it made of sub-units held together in some different way?

*c.* If such sub-units exist, what is the nature of the link between them?

As to the first question, the difficulty of accounting for the structure of spherical molecules out of a linear peptide chain has led to the idea of an alternate mode of linkage in which each amino-acid residue can be linked to four others and not merely to two. While this hypothesis at first sight has a theoretical attractiveness it still lacks any chemical support and has been subjected to serious criticism on chemical grounds.([7]) There are enough unknown factors in protein structure already without employing doubtful chemical assumptions. In retaining the peptide chain as the basis of protein structure it should, however, be remembered that for experimental reasons it is difficult to demonstrate in proteins and that therefore it is still to be accepted with a certain caution.

The question of the unitary nature of the protein molecule arises whether the inter-residue linkage is peptide or not. With a multiple link it is easy to construct models either of the cage or solid type, but as has already been pointed out it is difficult to reconcile such continuous structures with the definite 10A. discontinuities revealed by the X-ray analysis. With a peptide chain, however, the unitary solution becomes more difficult. It is difficult to imagine any kind of fold or coil by which a single chain can occupy the observable space and at the same time not be so intricate that its formation by any natural process would be enormously improbable. There is, however, much

evidence that at least the larger protein molecules
are not unitary in structure. In the first place,
some of them can be split in solution down to par-
ticles of molecular weight of the order of 10,000,
and this is probably not the lower limit as smaller
particles are difficult to isolate or measure. Two
lines of X-ray evidence also point to sub-units, the
high symmetry of protein crystals and the 10A.
repetitions. The symmetry of protein crystals is
much higher than would be expected statistically
from compounds of such great complexity. Cubic,
hexagonal and tetragonal crystals are found fairly
frequently among proteins, while very rarely among
even moderately complex organic compounds.
Even when the symmetry is lower there is a marked
tendency towards pseudo-symmetry. This sym-
metry can be accounted for in two ways, either by
some fundamental symmetrical element in the asso-
ciation of amino-acids, such as postulated in the
cyclol hypothesis of Wrinch, or by assuming that
the proteins are composed of sub-units which are
themselves unsymmetrical but arranged in a sym-
metrical way. The difficulties of the first hypo-
thesis have already been pointed out, whereas the
second not only fits in with what we know physi-
cally about the proteins but is also compatible with
the 10A. discontinuities. The size of the sub-units
must lie between that of the smallest protein
observed, that is one with a molecular weight 9,000
and that of a single amino-acid of molecular weight
averaging 120, that is it must contain some sub-
multiple of probably 72 amino-acid residues. The
uncertainty arises from the fact that it is not neces-
sary for all the sub-units to be the same, though
some must be to account for the symmetry. The
presence of trigonal symmetry suggests that the

assymmetrical unit must be a third or less of this number, that is it must contain 24, 12 or 8 amino-acid residues.

The question of the structure of the sub-unit would seem to raise at first sight the same difficulties as the structure of the molecule itself. Actually, however, with the smaller number of residues the difficulties and the improbabilities of the coiling become much less, particularly if we postulate—which is not unreasonable—that the sub-units are closed peptide rings. Such rings would necessarily curl up owing to the mutual attraction of the positive and negative charged amino and keto groups in the chain, and models of such coiled chains can be constructed which preserve distances between such groups which have been found in other compounds. The difficulty is, however, that there are a very large number of such possible models, and as yet nothing to choose between them. The method of folding in the chains in the sub-units may well be similar to that of the contracted form of the fibrous proteins mentioned above, and indeed a more detailed study of these may give us a clue to the whole arrangement.

The postulation of sub-units, however, raises further questions in the molecular structure, for they must be bound together sufficiently tightly to hold the molecule together in aqueous and ionic solutions. For this purpose a limited number of bonds are available. Ionic bonds are plainly out of the question, as they would certainly hydrate. There remains the possibility of amino-carboxy links between the ends of side-chains, but this is somewhat liable to the same criticism. It seems more probable that the links are either or both of two kinds, S-S linkages and association of hydro-

phobe groups. With the small number of sub-units the number of S-S linkages needed to hold the molecule together is, for all proteins hitherto examined, sufficiently provided for by the amount of sulphur present, and the extreme changes in activity that proteins undergo when S-S bonds are broken indicate that they may have this fundamental part to play. Whether this is so or not, however, the behaviour of the hydrophobe group of the protein must be such as to hold it together. As Danielli ([16]) and Langmuir([17]) have pointed out, on the basis of surface-film work, the protein molecule in solution must have its hydrophobe groups out of contact with water, that is in contact with each other, whereas on the surface the molecule is broken up into a film of 10A. thickness in which the hydrophobe groups are driven out of contact with the water. In this way a force of association is provided which is not so much that of attraction between hydrophobe groups, which is always weak, as that of repulsion of the groups out of the water medium.

Langmuir has used this picture as a justification of the cyclol cage hypothesis, but it is strictly quite independent of it, and the model outlined above has the advantage of accounting very satisfactorily for the phenomenon of denaturation, particularly on a surface. Once the sub-units come to the surface their rings are brought into a plane and different rings can interact according to the familiar ring-chain polymerization process, which will result in the formation of the fibres which Astbury has shown to exist in such films. This polymerization process takes measurable time as Danielli's work has shown.

The picture thus presented is far from being a

finished or even a satisfactory one. The crucial fact that requires elucidation is the precise mode of folding or coiling of the peptide chains, and for this we may have to wait for some considerable time until the X-ray and other methods have advanced their technique much further than at present. The problem of the protein structure is now a definite and not unattainable goal, but for success it requires a degree of collaboration between research workers which has not yet been reached. Most of the work on proteins at present is uncoordinated, different workers examine different proteins by different techniques, where a concentrated and planned attack would probably save much effort which is now wasted and lead to an immediate clarifying of the problem.

It is difficult to exaggerate the importance of this study to many branches of science. The protein is a key unit in biochemistry and physiology, and by studying its structure we are at the same time providing answers to a number of problems in these and wider fields. A fundamental question in protein structure is whether proteins are to be considered as necessary or contingent constructions. Is the protein molecule simply an arrangement of atoms or of amino-acids of such low energy that it will tend to be formed whenever suitable materials are found together, or is it, on the other hand, the product of long evolution of chemical forms owing its structure to a series of historical accidents in the development of life? The instability of proteins and the apparently arbitrary nature of many of their constituents speak against the first view, while if the second is accepted we have in the structure of proteins a key to the solution of many of the problems of the development and origin of life. All protein

molecules that we know now have been made by other protein molecules, and these in turn by others. The evolution of the protein molecule has run parallel with the evolution of the organisms which are constructed from them. The protein molecule may thus be considered as a living fossil. Certain of its characteristics, notably the presence of amino-acids of the same configuration, indicate a unity of pattern and point to some historic circumstance in the first production of protein molecules or their possible precursors. It is clearly too early to discuss these subjects in anything more than a speculative manner but the ideas they suggest lend an additional interest to the fascinating problem of the structure of proteins.

[J.D.B.]

## REFERENCES

1. W. T. Astbury, Nature, cxlii, 33 (1938).
2. J. Chem. Soc. (1935), 846.
3. T. Svedberg (Royal Society discussion on protein structure, 1938), in press.
4. J. D. Bernal and D. Crowfoot, Nature, cxxxiii, 794 (1934).
5. J. D. Bernal et al, Nature, cxli, 521 (1938).
6. M. Bergmann and C. Niemann, J. Biol. Chem. cxv, 77 (1936), cxviii, 301 (1937).
7. Neuberger (Royal Society discussion on protein structure 1938), in press.
8. I. Langmuir, J. Chem. Phys., vi, 873 (1938).
9. Levin, personal communication.
10. W. T. Astbury, "Fundamentals of fibre structure " Phil. Trans., A. ccxxxii, 333 (1933).
11. Biochemical Journal, xxix, 2351 (1935).
12. Personal communication.
13. Proc. Roy. Soc., A. clxiv, 580 (1938).
14. J. Amer. Chem. Soc., lx, 2247 (1938).
15. J. D. Bernal, Nature, cxliii, 74 (1939).
16. Danielli (Royal Society discussion on protein structure 1938), in press.
17. I. Langmuir and D. Wrinch, Nature, cxliii, 49 (1939).

# Friday, March 3, 1939

VISCOUNT FALMOUTH
*Manager and Vice-President, in the Chair*

SIR WILLIAM BRAGG, O.M., K.B.E., M.A.,
D.Sc., Pres.R.S., M.R.I.
*Fullerian Professor of Chemistry*

## LIQUID FILMS

THE study of the behaviour of liquid films has long been highly interesting, and in recent times very profitable to many sciences and to industries also. Several discourses have been given in this theatre by men who have made important contributions to the subject. In 1889 Sir Charles Vernon Boys described 'Soap-bubbles and the forces which mould them' to an audience of children in a series of Christmas lectures at the London Institution. His explanations and his remarkable experimental skill have made those lectures famous; the book which he wrote subsequently still preserves its first attraction.[*] Sir James Dewar in his later years took a deep interest in the properties of liquid films. He proved the extreme importance of purity and cleanliness, and was able to achieve results which were quite out of range before his time. Bubbles and films were made by him to last for months instead of minutes. His skill and experience have been handed on to some of those who worked with him, particularly to our lecture assistant and friend Mr. W. J. Green and to Dr. A. S. C. Lawrence who has written an excellent summary[†] of his experiences. The experiments of Dewar added new

[*]*Soap Bubbles, their Colours and the Forces which mould them.* S P C K 1912.

[†]*Soap Films.* G Bell & Sons, 1929

beauties and interests to the many that previous investigators had demonstrated and I have thought that it would interest the Members of the Royal Institution to see them once more after a lapse of fifteen years. Mr. Green will repeat some of them for us.

Moreover, new and important aspects of the film phenomena have come to light since Dewar's time. Members will remember the interest and significance of the results which Dr. Langmuir described here a few months ago, and the good work of Dr. N. K. Adam. We now have fresh information as to the form of the molecules of the soaps which play so great a part, and we have come to realize that we cannot form any clear picture of what is happening until we take molecular form into account. We have to reconsider the phenomena with this new knowledge of molecular structure to help us; this is a point on which Lawrence has rightly insisted.

If we are to see some of the experimental aspects this evening, we must be content to spend little time on history and theory. I will refer only by name to pioneers such as Plateau, Perrin, Wells, Johannot, Reynolds, Rücker, and even then must be satisfied with a very short list. I must go on at once to the relation of the chief points to be observed and to the consideration of such explanations as can now be given. We have at our disposal some of the pure ammonium oleate which Dewar found so useful. An ordinary soap solution of the commonest kind will yield bubbles and films but they do not last and their phenomena are complicated by reason of impurities and dirt. Many of the most interesting effects do not appear at all unless the materials are pure and the apparatus perfectly clean. The liquid from which our films are
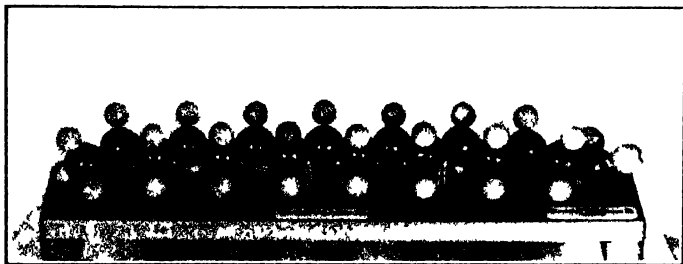
**A** A single molecule of Palmitic Acid   The black balls represent the carbon atoms, the small white ones hydrogen atoms, and the larger white ball at the right hand end an oxygen atom
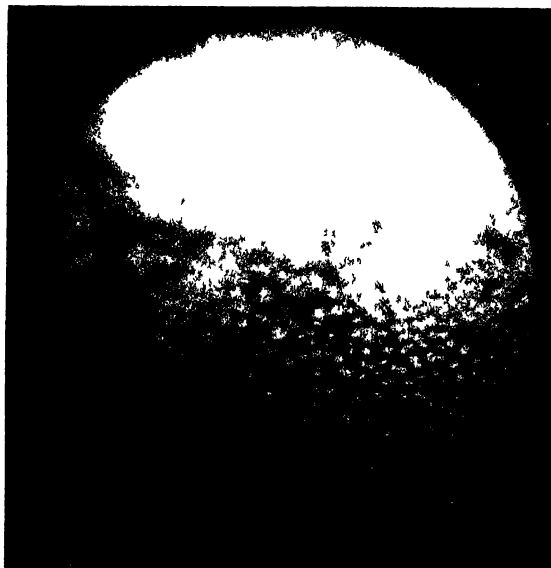


**B** Microscopic appearance of a thinning soap film Minute droplets are extruded and are seen coalescing on the film surface, from whence they drain away to the boundaries   The actual area shown is less than 2mm diameter

PLATE I.

formed is a mixture of water, a salt of oleic acid and glycerine. The form of the oleic acid molecule is of the first importance; the chemical formula is $CH_3 (CH_2)_7 CH = CH (CH_2)_7 COOH$.

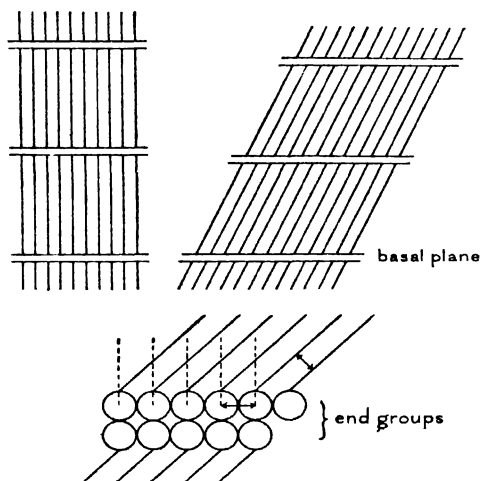The arrangement of the atoms in a long-chain



FIG. 1. Diagrammatic sketch of the arrangement of long chain molecules. The upright model is that of a paraffin crystal, the sloping one represents a fatty acid crystal and the lower diagram suggests the kind of arrangement where opposite ends of the fatty acid molecule link together.

structure of this kind has been determined by X-ray crystal analysis. For example, Plate IA shows a model of the structure of palmitic acid. The carbon atoms are arranged in a zig-zag order, successive links making with each other the tetrahedral angle (109° 28'). The hydrogens are represented by small balls and the two oxygen atoms at one end by balls of medium size. The distance between the centres

of two neighbouring carbon atoms is 1.54 Ångstrom
units (of which a centimetre contains a hundred
million). Oleic acid differs from palmitic in that it
has two more carbon links in the chain, and also that
two of the carbon atoms in the middle of the chain
have only one hydrogen atom each instead of two.
There is in consequence a double bond between
these two carbon atoms. This is no doubt very
important and gives to the oleic molecule special
film-making properties, though the reason is not
clear. Any single bond can serve as an axis round
which one part of the molecule can be rotated with
respect to the other. Thus the molecule can be
coiled up. In the solid state (the longer the chain
the higher the melting point) the molecules lie side
by side, as represented in Fig. 1. The tendency to
lie side by side must be due to the van der Waals
forces exerted between the side atoms of the one
molecule and those of the next. The crystal structure
of oleic acid in the solid form has not been deter-
mined but there seems no reason to doubt that the
same tendency for the molecules to lie side by side
exists. It has been proved for stearic and for
stearolic, which resemble oleic in all respects as
regards composition, except that in stearic every
carbon atom in the chain carries two hydrogens
and in stearolic two carbon atoms in the centre are
without hydrogens. In oleic the two have one each,
placed on the same side of the molecule. It is
curious that elaidic acid, which is like oleic except
that the two hydrogens in the centre are on oppo-
site sides, is useless for our present purpose. Films
of elaidic are streaky and inhomogeneous. Full in-
structions for preparing oleic acid and ammonium
oleate from oleic acid are given by Lawrence.

Langmuir, Adam and others have shown that

the films formed of fatty acids and other similar substances by their spreading on the surface of water consist of molecules standing on the water side by side. The carboxyl end of the acid, or the substitute for it containing an alkali, is rooted in the water for which it exercises an attraction, and the film exposes only methyl terminals to the air above. The film is like a hair pile.

It has long been supposed that such a pile forms the face on each side of a soap film. When the film is first formed or the soap bubble first blown, the space between the two faces contains liquid in such quantity that there are no interference colours. As the film drains thin, colours appear in a regular succession, known since Newton's time. The linking of the long molecules side by side on each surface forms a containing sheet; just so lines of soldiers keep back a crowd. The formation of the sheet is not instantaneous. In any ordinary mass of the solution the soap molecules are distributed at random, they begin to assemble on the surface because, to put it broadly, the water molecules have a greater attraction for each other than for the soap molecules, and therefore tend to expel the latter in the course of time. Rayleigh proved this gradual approach to a final state; the surface assemblage was incomplete on the surface of a jet rising rapidly from an orifice. He found, however, that the final state was quickly reached.

When any two liquids are mixed molecularly there must always be a tendency for one to gather on the surface in greater molecular numbers than the other; the one in which there is the strongest affinity of neighbour to neighbour will tend to expel the other. Yet owing to thermal motions the interior will never contain molecules of one kind

only while all the others lie on the surface. There will be only an excessive concentration of one kind on the surface and of the other in the interior. The excesses will depend directly on the energies spent in overcoming the relative affinities. Interchanges between molecules on the surface and inside take place continuously.

This is a simple problem but the problem of the soap film is more complicated  When the molecules of the soap arrive at the exterior they do more than accumulate, (some always coming some always going) they arrange themselves in an incipient crystallization, and this introduces an entirely new feature. It is this which gives strength and permanency to soap films. It is a factor which is not taken into account in the simple surface tension formulae. Not merely the number of molecules per unit volume or per unit area of surface but the mode of arrangement and the forces consequent on that arrangement become important.

The outer surface of a bubble is a sheet of methyl groups and slides very easily over a similar sheet. In the crystal—of paraffin or any other such waxy and greasy material—the methyl sheets form the slip planes. One soap bubble can therefore be rubbed over another or pressed against it without forcing the two to coalesce or to burst. Boys showed the children some beautiful experiments based on this fact; one bubble forced another through a ring; one could be blown inside another and rest against its inner surface without causing a collapse, or be filled with hydrogen and carry the outer bubble upwards.

Films and bubbles always drain to a less thickness. If the film is vertical gravity will be one cause of the effect but there is in general a second

cause. There is a ring of liquid all round a film
bordering on the frame that carries it. This may
be understood from simple considerations. Suppose
that the Figure 2a represents a section of a box
or pipe in which a certain quantity of soap solution
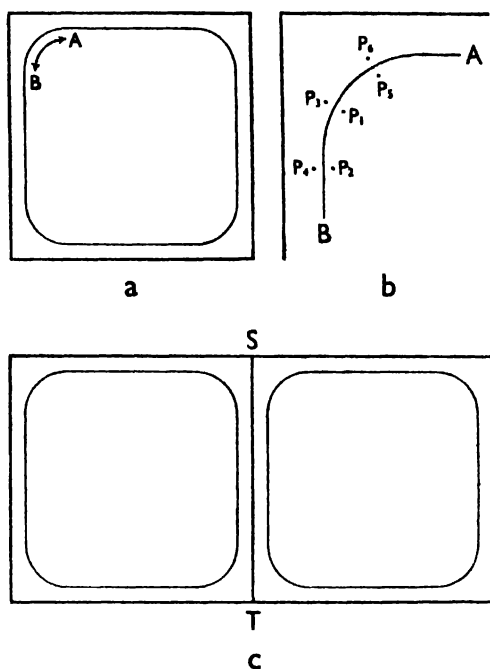


a          b

S

T

c

Fig. 2

has been placed. Let the box be turned over a few
times. In the end there will be a film of solution
all over the walls and a collection in each corner.
The section A to B, and at the other corners, is a
circle, since at any two points on the inside of the
bend the pressures $P_3$ and $P_6$ (Fig. 2b), are equal,
and so are the pressures at $P_1$ and $P_5$. Now the
difference between $P_1$ and $P_3$ and between $P_5$ and

$P_6$ depends on the curvature at the two places respectively. Since the differences are equal so are the curvatures and the section is circular.

But at a point on the side where the film is flat, $P_2$ is equal to $P_4$, and since $P_2$ is equal to $P_1$, there must be a difference between $P_4$ and $P_6$, the former being the greater. Consequently the liquid tends to drain away from the sides into the corners.

Suppose we had a double box as in Fig. 2c, and that the films in the two are alike, we can imagine the material sheet at ST to be withdrawn, leaving an ordinary liquid film across the box. The film will gradually thin as the liquid leaves it for the channels at S and T. It will thin most quickly in the middle.

The walls of the boxes will never dry completely if there is, as usual, a strong attraction between the liquid and the walls. Neither will the film at ST ever thin to nothing in the centre, because when the two surfaces come very close to one another the two sets of molecules on the two sides lock into one another and form a sheet of some permanence, the well known *black spot*, black because it is so thin that it cannot reflect any interference colours. Its thickness is about 50 Ångstrom units, about twice the length of the chain of oleic acid when straight, as the chains of stearic acid are straight in Fig. 1. Probably a few molecules of water or glycerine are concerned in the formation of the two sides and allow the molecules to be perpendicular to the film, not sloping as in the crystal, but this is little more than conjecture.

The black film may take hours or days to spread completely over a bubble or film. Many records of observed times under varying conditions are to be found in Dewar's papers. The boundary between

the black and the neighbouring part of the film is extraordinarily sharp. The one does not gradually melt into the other as in Fig. 3A, but there must be a steep bank at 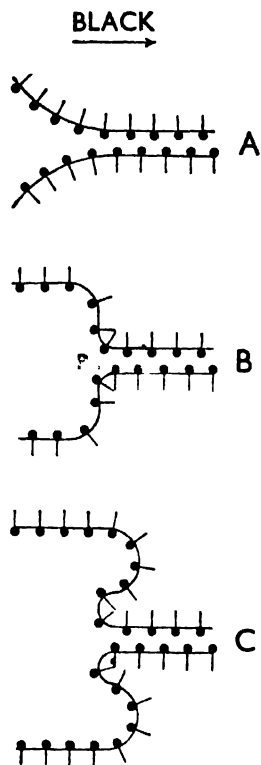the boundary as in Fig. 3B. If we remember that the conditions are not stable and that the black is slowly encroaching we may find an explanation of this remarkable effect. At P the two layers of molecules are beginning to interlock. On the right of P the interlocking is complete, in a fashion that reminds one of a *zip* fastening. As the interlocking at P proceeds, molecule by molecule, the material to the left of P has to be squeezed away from P, and this takes time. If it cannot get away quickly enough, there must be a bank at P.

It does not seem unlikely that the interlocking process will cut into the bank as in Fig. 3C and start the formation of a block, which is kept in contact with the film by the van der Waals forces, and yet is very free to slide over its surface. The films often seem to be extruding liquid, and this may well be the way in which they do it. Blocks are observed sliding about on the surface of the black film. In an upright film such blocks gravitate downwards. This may occur in an early stage of thinning (Plate I B) or later when stratification has become more regular

BLACK

A

B

P

C

FIG. 3
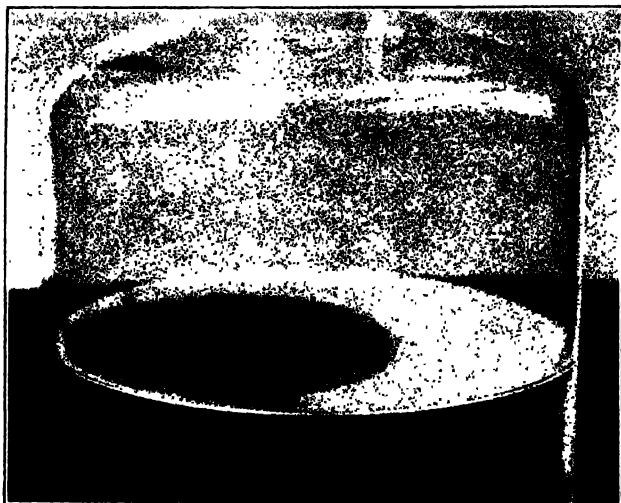
A. Graded black areas, each has a definite boundary. These often coalesce individually into silvery droplets. Similar definite boundaries are often developed between coloured zones in a soap film kept for some time under quiet conditions.



B. Shower of silvery droplets over a black film in vacuum.

PLATE II.

and there are steps in thickness from black to grey (Plate II A). The appearance is that of black spaces moving upwards.

In one of Dewar's beautiful experiments an upright film is formed in a vacuum, under which conditions a black film is soon formed. When a little heat is applied to the glass containing-vessel close to the lowest point of the film, some of the liquid which is assembled there runs upwards round the two sides of the circle, and the two streams meet at the top, whence they descend in a shower clinging to the sides of the film. The illustration (Plate II B) shows this effect. The shower consists of small slabs of uniform thickness, which slide downwards over the black surface.

The formation of sheets or slabs, or *plages* to use the term employed by Perrin and other French workers, is a very obvious phenomenon. Perrin observed that the process was greatly promoted if a quantity of uranin or other fluorescent substance was introduced into the solution and illuminated by that particular light which excited the fluorescence. It would appear that in some way the stratification is encouraged by the presence of ions or free electrons, probably because in this way the re-arrangement of molecules in semi-crystallization is made more easy. It may well be that the arrangement of the oleic acid molecules on the surface is accompanied by an arrangement of similar molecules and water molecules below the actual surface, and that this partial ordering favours the layer formation characteristic of long chain molecules. [W.H.B.]

During the discourse a number of experiments due to Dewar and others were repeated and shown to be in accord with the recently developed views respecting molecular form and arrangement.

# Friday, May 5, 1939

R. S. WHIPPLE, M.I E.E., F.Inst.P.
*Manager and Vice-President, in the Chair*

W. L. BRAGG, O.B.E., M.C., D.Sc., F.R.S., M.R.I.
*Professor of Natural Philosophy; Cavendish Professor of
Experimental Physics in the University of Cambridge*

## MAGNETS

[ABSTRACT]

IT has long been known that steel, when magne-
tized by placing it in a magnetic field, remains a
permanent magnet when the field is removed,
whereas soft iron under the same conditions loses
practically all its magnetization. Before 1910 glass-
hard carbon steel was practically the only material
used for permanent magnets, but since that time
various new alloys have been discovered which
have properties very greatly superior to those of
carbon steel. Tungsten steel, containing 0.5 per
cent. carbon and 5 per cent. tungsten, was one of
the first of these new alloys. It was followed by
chromium steel and by a 35 per cent. cobalt steel
which is still largely used. In 1934 a completely
new field was opened by Mishina in Japan, who
found that an alloy of approximately two parts of
iron to one of nickel and one of aluminium was very
much better than anything yet discovered. For
instance, a small magnet of this type in which
the active material weighs about two ounces* is
capable of supporting a weight of 56 pounds. It
is very interesting to investigate the origin of these
remarkable magnetic properties.

*Shown at the lecture by the courtesy of Mr. D. A Oliver of the
Permanent Magnet Association.

In measuring the effectiveness of a permanent magnet material, one must distinguish between "remanence" and "coercive force". Suppose the material in the form of a closed ring to be magnetized by a current in a coil wound round it. When the field is reduced to zero by cutting off the current, all ferromagnetics retain a large proportion of their magnetization, to an extent measured by the "remanence". This property by itself is no measure of their usefulness as magnets, because in this "zero field" condition the magnetic lines of force are in the closed circuit of the ring and do not produce a field outside the material. If now a sufficiently high reverse field is applied by reversing the current, the magnetization is reduced to zero. The "coercive force" is defined as the strength of the field which destroys the magnetization. It is small for soft iron, larger for steel and very much greater for the new permanent magnet materials. It is the coercive force which enables a bar magnet to retain its magnetization against the effect of its own poles, and so to store up magnetic energy and produce an external magnetic field. The effectiveness of materials for permanent magnets is roughly proportional to the product of remanence and coercive force (actually the magnetic energy which can be stored up in them is measured by the maximum value of the product BH, H being the reverse field and B the remaining induction). It is not possible to increase the remanence greatly, as it must in any case be less than B for saturation, and this is determined by atomic constitution. Actually the remanence is rather poor for the new magnetic materials; the improvement is in the value of the coercive force, which is about ten times as great for the new materials as for tungsten steel. The effect of this

change is that whereas former magnets had to be long and thin in order to retain their magnetization against the reverse field of their poles, the new magnets can be short and thick. A given field-strength over a given volume can be achieved with a much smaller weight of material.

Two conditions must be satisfied for a material to be ferromagnetic. Its atoms must be elementary magnets and the interaction between them must be such as to make neighbours point in the same direction. An "unmagnetized" mass is divided into regions called "domains," each of which is completely magnetized because all its atomic magnets point the same way, but neighbouring domains are orientated in different directions so that on the average there is no external field. The size of these domains can be estimated by the Barkhausen effect. As the material is magnetized by an increasing field, the domains reverse individually and point in the direction of the field. A coil is wound round the specimen and connected to an amplifying system and loud speaker; a click is heard as each domain reverses. On the average the domains contain about $10^{15}$ atoms, or to put it alternatively, there would be about 1,000 domains in a pin's head. In a material with low coercive force, the direction of the domains is reversed by a very small field, whereas for high coercive force a strong field is necessary.

It appears that for a high value of the coercive force two factors must co-operate. In the first place, the magnetostriction coefficient must be large. Magnetostriction is the alteration in dimensions of a ferromagnetic when magnetized. For instance, iron expands for magnetization along a cubic axis (positive magnetostriction), nickel and

cobalt contract for all directions (negative magneto-striction). Conversely, if a material with positive magnetostriction is stretched, the domains all tend to point along the direction of strain, and if com-pressed, across it. In the second place, the material must have intense local strains. When there is a high magnetostriction coefficient and internal strain, the directions of magnetization of the domains are determined by the strains. Becker has made a simple model of what happens on magneti-zation. A low field alters the domain boundaries slightly, but the field has to be so strong that it overcomes the directional effect of strain before the domains can all turn over in the direction of mag-netization. Conversely, once turned, it takes a strong field to reverse them.

The high coercive force of the new permanent magnet alloys appears to be due to their intense internal strains. Bradley has recently explained the origin of these strains, using X-ray analysis to explore the atomic arrangement in the crystals. In the iron-nickel-aluminium alloy the atoms are randomly arranged on a body-centred cubic lattice at high temperatures. If the alloy is cooled very slowly this splits up into separate body-centred crystals of almost pure iron and a nickel-alu-minium-iron alloy with relatively little iron; in this state the magnetic properties are not exceptional. But if the cooling is at a certain optimum rate, partial segregation of the constituents takes place without separation into separate crystal lattices. As the iron-rich constituent has a smaller volume per atom than the other constituent, its coherence to a joint lattice-scheme throws it into a state of intense strain. This explanation is confirmed by a study of the iron-copper-nickel alloys which dis-

play a similar partial separation and also have a very high coercive force.

In contrast to the above, alloys of the permalloy type which are magnetized by extremely weak fields have a low magnetostriction coefficient and are annealed to remove as much strain as possible.

The new permanent magnet alloys were discovered by empirical methods, but now that the origins of their exceptional properties have been traced it may be possible to explore more systematically and to find materials with even more striking properties.

[W.L.B.]

# NEW PHENOMENA IN LIQUID HELIUM

LIQUID helium has long been recognized as a peculiar liquid. One of its most remarkable properties is that it remains a liquid however much the temperature is lowered and is likely to remain so down to the absolute zero. This unique property results from the quantum phenomenon whereby the helium atoms in the liquid at absolute zero retain a zero point energy of oscillation which is so large that solidification cannot occur.

It has more recently been discovered however that there are two quite distinct kinds of liquid helium with completely different properties. As the temperature of the liquid is lowered from 4.2 degrees to 2.2 degrees it behaves in every way like a perfectly normal liquid. Below the temperature of 2.19 degrees however its properties change completely. Thus it was discovered by Kamerlingh Onnes that whereas above 2.19°K. the liquid expands in a normal way on heating; below 2.19°K. it contracts on heating. It was found also by Keesom and his co-workers that the specific heat of the liquid rises suddenly below 2.19° and then gradually falls off again. This sudden increase in the heat required to raise the temperature of the liquid through a fixed amount is very similar to that which occurs when an alloy is changing from

a disordered to an ordered state or when a ferro-
magnetic body is changing from the state of order
characteristic of ferromagnetism to that of com-
parative disorder corresponding to paramagnetism.
It is therefore evidence for some change of state
whereby the liquid becomes more ordered as it
passes below 2.19°K. This phase of the liquid is
now known as liquid helium II. Further evidence
that liquid helium II is an ordered state is provided
if we compare the entropy curves of liquid helium
II and helium solidified under pressure, for we
then find that the entropies of solid and liquid
helium become equal as the temperature is lowered.
Now entropy is a measure of order and a solid
state is essentially an ordered state as compared
with a liquid —we infer then that liquid helium II
is a liquid in some special kind of order.

The more spectacular properties of liquid helium
have however been discovered within the last two
years. It was noticed at an early date that at tem-
peratures below 2.19°K. the liquid suddenly
becomes quiet and bubbles cease to rise. This can
be seen in the two flasks of helium I have here.
The reason for this was not discovered however
until two years ago when Professor Keesom made
an experiment to find the conductivity for heat of
the liquid. He set out to do this using the methods
suitable for a normal liquid which conducts heat
badly. He passed a known amount of heat through
a thin slab of the liquid and measured the tempera-
ture difference across it. To his surprise he was
unable to measure any temperature difference,
showing that the liquid conducted heat very well
indeed. He had therefore to proceed to the oppo-
site extreme and to make the heat flow through a
long fine capillary of liquid helium so that its

resistance to heat flow should be increased as much as possible. Even in this case it proved difficult to measure the temperature difference and it turned out that liquid helium II conducts heat 200 times better than copper or any other known substance.

This is not the whole of the story, however. Experiments were carried out in the Mond Laboratory by Allen and others to try to measure this high thermal conductivity more accurately, using a new and more sensitive method illustrated in Plate IA. The main difficulty in the experiment is to measure the extremely small temperature differences produced by the heat flow. In this apparatus the column of liquid helium used for conducting the heat is used also as a pressure gauge. A temperature difference at the two ends of the column produces a difference in vapour pressure of the helium which in turn produces a difference of level in the manometer. In this way temperature differences of a hundred thousandth of a degree were observable. The remarkable result then emerged that the conductivity for heat was not a constant as for all ordinary substances but is extremely high when small quantities of heat flow and falls off for larger heat flows. The thermal conductivity also increases rapidly with decreasing temperature, reaches a maximum at about 2.1° and then falls off again.

This apparatus, however, seemed to misbehave itself badly at low temperatures, for it was found that in some cases when current was passed through the heater the level of the liquid rose, whereas it would always be expected to fall with increasing vapour pressure above it. Fortunately Dr. Allen did not leave this peculiar behaviour without further examination and he constructed next the following
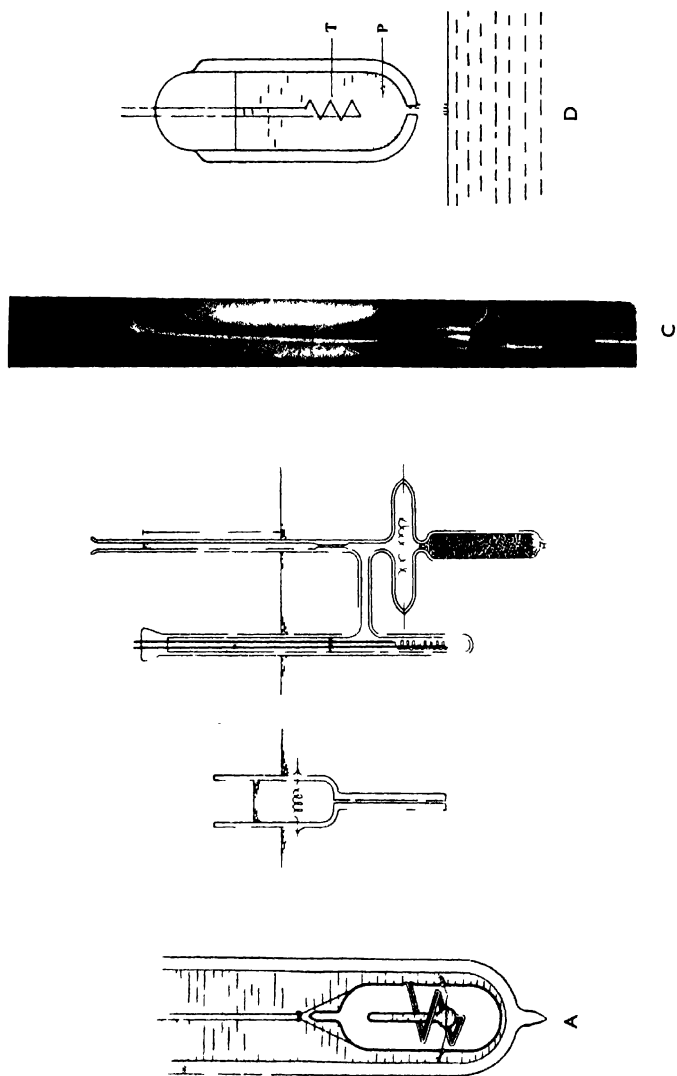
PLATE I

simple piece of apparatus (Plate 1B). A glass tube with a small heating coil is placed in a vessel of liquid helium and current is passed through the coil. It was then observed that the liquid level rose in the tube, showing that as heat flows down the tube a force acts on the liquid in the opposite direction causing its level to rise. This force must come from the wall or heater and so an equal and opposite reaction must be exerted on the wall. It follows, therefore, that the flow of heat down the tube is associated with a flow of momentum, some of which is transferred to the walls. Such a phenomenon associated with heat flow has never previously been observed, for ordinary elastic waves by which heat is normally conveyed do *not* convey momentum.

It seems also that the process is reversible for if we cool the liquid instead of heating it the level falls. This can be demonstrated in a simple way by using as a powder a paramagnetic salt which can be cooled by demagnetization.

The phenomenon becomes even more striking if the heat is made to flow downwards through a tube filled with fine emery powder. In this case the momentum transferred to the liquid becomes so large that a jet of liquid helium (Plate 1C) about a foot in height can be produced. This direct conversion of heat into mechanical work is certainly the simplest heat engine which has so far been discovered. It was therefore of special interest to measure the efficiency of the process and to see how this compared with the ideal engine working between the same limits of temperature. The results showed that the efficiency increases with decreasing temperature, reaching about one-tenth of the ideal efficiency. It is not likely that the

process can be of any practical importance but it does provide a very simple circulating pump for liquid helium.

Still another striking property of liquid helium II is found in the behaviour of its viscosity. It had been shown some time ago in Toronto by Misener and Burton, using an oscillating cylinder method, that the viscosity of liquid helium drops by a factor of 10 as it passes from liquid helium I to liquid helium II. When these new results on the properties of liquid helium II were found, Misener working in the Mond Laboratory with Allen took up the investigation again and measured the rate at which liquid helium II flows through very fine capillary tubes about a metre long. Thermometer capillaries of a few thousandths of a cm. diameter were used and it was at once found that the liquid flowed through these tubes 5,000 times as rapidly as the least viscous gas would have done, and a million times as rapidly as the most fluid liquid. The phenomena became even more striking as the capillary diameter was reduced. It is difficult to obtain glass capillaries below about 1/1000 cm. in diameter, but Dr. Allen succeeded in making extremely fine capillary passages by packing about 1,000 stainless steel wires into a german silver tube and then drawing the whole assembly through a series of dies. In this way the wires are squeezed more and more tightly together and very fine capillary channels of about $10^{-5}$ cm. width are produced. For such capillaries the average flow velocities reach about 12 cm. per second with a pressure head of 100 dynes whereas if hydrogen, gas under the same pressure were passed through its velocity would be only $10^{-5}$ cm. per second. Another striking result is that the flow becomes

almost completely independent of the pressure
head across the tube. In general, therefore, the
finer and longer the capillary the more queerly it
behaves. If we examine the flow in short, large
diameter tubes, the liquid behaves almost normally,
so much so that if the experiments are limited to
short capillaries one may be led to believe that there
is nothing abnormal in its properties. In order to
obtain a check on these results the flow of the liquid
through very fine powder was measured. Jewellers'
rouge was packed tightly into a glass tube 10 cm.
long and 3 mm. bore. Helium gas was then driven
through the powder under 10 atmospheres pressure
to remove the air and at the same time the volume
of gas flowing was measured and found to be about
$10^{-4}$ c.c. per second. The flow of liquid helium
through the powder was then determined at differ-
ent temperatures under a pressure difference of
$10^{-4}$ atmospheres. Just as in the capillary flow
experiments the velocity increased rapidly with
decreasing temperature until at a temperature of
1.2 degrees about 1/20 c.c. per second was flow-
ing. In these experiments again, then, the flow is
about a million times greater than with helium gas
under comparable conditions. The flow through
powder behaves differently from flow through fine
capillaries, however, in that the flow is propor-
tional to pressure for the powder and is almost
independent of pressure for the capillaries.

A third phenomenon which may ultimately pro-
vide the key to all was discovered by Kamerlingh
Onnes 16 years ago. Onnes discovered that if two
vessels containing liquid helium are connected,
the level of the liquid in the two tubes shows a
mysterious tendency to remain the same. If one
of the vessels is emptied, liquid is transferred from

the other till the levels are equalized. This phe-
nomenon appeared to have excited little attention
until the recent work on the liquid. Mendelssohn
then took up the investigation of this transfer
effect in greater detail. If an empty test tube is
dipped into a flask of liquid helium II the test tube
slowly fills, as though there were a hole in the
bottom. Just as in the case of flow through fine
capillaries the rate of flow remains almost inde-
pendent of the difference of level between the two
vessels until equality of level is reached, when the
flow stops. If now the level of the test tube is
raised liquid flows out of the tube until the levels
are again equal. If we assume that there are really
no holes in the test tube, only two possibilities are
open. The liquid can be transferred by evapora-
tion and recondensation or it can move over the
surface of the glass in a thin film. A decision
between these two possibilities was reached by
increasing the surface connecting the two vessels
by copper wires. It was found now that the rate
of transfer was increased proportionally to the
increased area of the surface. It seems, therefore,
that surfaces in contact with liquid helium II
are covered with a thin film of the liquid moving
with an appreciable velocity. An estimate of the
thickness of the film was made by Mendelssohn by
allowing the film to form on a surface of known
area. The surface could be wound up into a warmer
part of the Dewar flask when the film is driven off
and forms a liquid in the bottom. In this way the
thickness was found to be about $4 \times 10^{-6}$ cm. The
rate of transport of liquid was now investigated
and found to increase rapidly with decreasing tem-
perature reaching a maximum of $0.75 \times 10^{-4}$ c.c.
per cm. width of surface per second at $1.5°$. From

this the maximum flow velocity is found to be 20 cm. per second.

The existence of this film had actually been suggested earlier by Simon and Rollin as a result of some of their experiences in working with liquid helium. In all such experiments one is anxious to reduce as much as possible the rate at which the liquid helium evaporates from the apparatus. They found that the rate of evaporation increased rapidly below the transition point. This increase in the evaporation must have been caused by an additional heat loss from the vessel and they suggested that it might be caused by the existence of a thin film of liquid covering the walls of the vessel. The liquid had just recently been found to have a high thermal conductivity suggesting that such a film might conduct heat from the bath of liquid up the walls of the tube. Alternatively the liquid film might be considered to creep continuously up the walls of the tube and to evaporate at the top where the temperature is high. There would thus be a continued circulation of liquid conveying heat out of the liquid to the walls of the vessel. From Mendelssohn's work on the transfer effect it seems likely that the second mechanism is mainly responsible for the heat loss but the first may also play some part.

The existence of the film suggests possible explanations of the other peculiar phenomena in liquid helium. Thus the characteristics of the flow through fine capillaries shows a marked resemblance to those of the surface flow. Both set in at the λ point and increase with decreasing temperature in much the same sort of way, whilst the average velocity in the finest capillaries seems approximately equal to the film velocity at the same temperature.

We may therefore assume that part of the flow in capillaries is due to a mobile surface film. It is then reasonable to expect that this mobile film prevents the normal retardation of the remaining fluid by the walls so that it is dragged along with an average velocity comparable with, though not equal to, the film velocity.

This mechanism explains also the observation that increasing the length of a capillary tube does not decrease the velocity by as much as would be expected classically. For the film velocity is almost independent of length of the surface so that the surface component of the flow may hardly change with increasing length.

Strong support for this explanation has been provided by further experiments at Cambridge. When liquid helium flows through fine capillaries into a helium bath the flow is observed to overshoot itself and oscillations of the liquid level follow, and it is a simple measure to calculate from this the inertia and so the cross section of the column of moving fluid. If, however, we replace the capillary tubes by a bundle of the fine wires used to make the capillaries and observe the transfer over the surface of these wires, Allen finds exactly the same type of overshoot and oscillation. This suggests strongly that the mechanism of flow is similar in the two cases.

A further confirmation has been provided by the discovery that the fountain effect or transfer of fluid produced by heat flow has its counterpart in the film phenomena. Thus if a test tube is dipped into a vessel of liquid helium II and the levels allowed to equalize and heat is then supplied to the test tube by a heating coil suspended by a thread, it is found that the level of liquid in the test tube

rises by as much as 5 millimetres, flow along the surface being induced by the heat flow. Now we have seen that the fountain effect is increased in powder tubes where the surface is large. All this suggests that the explanation of the fountain effect must be linked with the surface flow phenomena.

It would seem also that the violent circulation of liquid produced by the fountain effect may be largely responsible for the high conductivity for heat shown by the liquid. Since in this case heat is carried by mass transport, the heat conductivity should be largely determined by the specific heat of the liquid. This in fact is found to be the case, for with decreasing temperature the conductivity for heat rises rapidly to a maximum in much the same way as does the specific heat and thereafter both tail again to quite low values at temperatures of the order of $0.5°K$. It has still to be shown, however, that the explanation fits quantitatively as well as qualitatively. If we may assume this for the time being it will be seen that most of the peculiarities of liquid helium might be explained if we can explain the film phenomena.

Two possible explanations of the phenomena in liquid helium II have been put forward. The first, which is now rather out of favour, assumed that the liquid helium II has an ordered structure in space—that it is a crystalline liquid. On this picture the large specific heat would be accounted for by a change from order to disorder as the liquid warmed up, whilst a lattice with tunnels running through it was invoked to explain the low viscosity. This hypothesis was put to the test by Keesom and Taconis who investigated its structure by X-rays. A narrow pencil of X-rays was passed through a horizontal jet of liquid helium produced by a small pump and

a semi-cylindrical film placed to receive the scattered X-rays. If the liquid possessed some kind of crystalline structure characteristic spots might have been observed if the jet behaved like a rotating crystal, or a series of curved lines would, have been found if the jet behaved like a series of small crystals. Instead of this, however, Keesom found only a ring characteristic of a liquid structure. These experiments have been repeated in Cambridge by Mr. Reekie at different temperatures and again only evidence of a liquid structure has been found.

An alternative explanation of the phenomena put forward by London and Tisza is that we have here an example of the so-called "degeneracy phenomenon" of Quantum Mechanics. It was long ago shown by Einstein that at very low temperatures a gas such as helium should show peculiar properties. As the temperature is lowered below a critical value, which should be in the region of three degrees, it turns out that some of the atoms *must* condense into a state of zero energy and that the lower the temperature the more of them are in this state. London assumes that the Einstein condensation phenomenon is actually occurring in liquid helium but that the closer spacing of atoms in the liquid distorts the phenomenon and reduces the critical temperature from 3.1 degrees to 2.19 degrees. He also makes the reasonable assumption that the condensation occurs more readily near the walls of the vessel owing to the larger forces between the wall and a helium atom than between two helium atoms. In this way he identifies the film with a high density of condensed atoms. If we now inquire what properties we might expect in a film which is largely composed of atoms of zero energy, we

might observe first that the zero energy of these atoms must mean that the wave function defining their properties must spread out over the whole vessel and that these atoms are so diffuse that they might behave almost as though they were not there at all. The results would be that the remainder of the atoms—the uncondensed ones—having little interaction with the condensed atoms, are effectively well spaced.

We might also expect that when atoms condense into the state of zero energy heat is liberated and conversely is absorbed when they come out of that state. It was therefore possible to produce a plausible explanation of the fountain effect, by assuming that as heat is given to the liquid at one point in a tube, atoms come out of the condensed layer in the wall to absorb the heat and that a motion of the fluid takes place along the wall to replace those "evaporating". We have therefore a cyclic process in which atoms condense at the cold end, run along the surface and evaporate at the hot end, producing a surface flow of liquid in a direction opposite to the flow of heat. This cyclic flow must then be responsible also for the heat conductivity.

The maximum efficiency of the fountain effect as a method of transferring heat into mechanical work can easily be calculated by considering the process as a heat engine. For if $q$ c.c. of the liquid flow per second against a pressure difference $dp$ then the efficiency of the process is

$$\frac{q \, dp}{H}$$

This must be less than $\frac{\Delta T}{T}$ where $\Delta T$ is the temperature difference between the hot and cold ends.

Actually as we have seen, the efficiency reaches only about 12 per cent. of the ideal efficiency.

We may also test the hypothesis by reversing the heat engine. For if heat produces flow then flow might produce heat. Thus if the flow out of a vessel takes place largely by surface flow, then as the atoms condense into the surface film they will liberate heat and the temperature should rise.

This effect was sought for in smooth capillaries by Dr. Allen but proved to be too small for detection. Recently, however, Mendelssohn has succeeded in detecting a temperature rise of 1/100 degree as liquid helium flowed out of a vacuum vessel through a fine powder (Plate 1D). Similar experiments were performed by Dr. Allen but in his case the temperature rise did not exceed 1/2000 degree.

It is interesting at this point to calculate the order of magnitude of the temperature rise to be expected. In Dr. Allen's experiment, about 1/10th of a c.c. of liquid was flowing out per second, doing mechanical work at the rate of 15 ergs per second. If there were simply a direct transformation into heat then the rate of rise of temperature in the vessel could not exceed $2 \times 10^{-7}$ degrees per second. But if the process acts like a reversed heat engine— a heat pump—then it is well known that much more heat can be pumped than corresponds to the work done. And so we can see why the much larger temperature rises are produced.

But these experiments are full of pitfalls and I should not yet like to regard it as proved that these phenomena can be described in such simple pictures.

For there are many difficulties. First, as to the nature of the film itself. The film is observed to

spread over all the surfaces in the neighbourhood of the liquid bath and causes these surfaces to be anchored rigidly to the bath temperature. The film therefore must have the bath temperature and yet is supposed to consist largely of "condensed atoms."

Second, it seems surprising that the velocities of flow in such films should not reach molecular velocities. Recently in Cambridge Dr. Ganz has been trying to measure the velocity of a heat impulse in liquid helium and finds that it travels with a speed of at least 100 metres per second. It is possible, however, that this first impulse may be transmitted in some other way.

Third, although the film velocities and the flow in fine capillaries seem almost independent of pressure heads and temperature differences producing the flow and reach maximum velocities of about 20 cm. per second, it is clear that much greater velocities are required to explain the observed heat transfer in smooth bore tubes.

[J.D.C.]

# PHYSICAL SCIENCES

The discourses at the Royal Institution are unique in character. On each Friday between October and June, a well-known authority is invited to give a general review of his subject, and he is asked not to assume any expert scientific knowledge on the part of his listeners. A tradition has grown up around the "Royal Institution Discourse" which is regarded as a special occasion calling for a talk distinguished by simplicity and clarity and, when appropriate, illustrated by interesting experiments and demonstrations. Most of the discourses are on scientific subjects, but the arts are also represented.

These talks present a broad survey of current scientific thought and achievement. They carry the authority of the famous men who have given them, and at the same time they are easy to understand. Their great interest provides the justification for their being published in the present form. No attempt has been made to select; all have been included because it is rare to find one lacking in interest today.

## APPLIED SCIENCE